(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 June 2005 (16.06.2005)

PCT

(10) International Publication Number
WO 2005/054461 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number:
PCT/US2004/039805

(22) International Filing Date:
23 November 2004 (23.11.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/525,318	26 November 2003 (26.11.2003)	US
60/542,520	6 February 2004 (06.02.2004)	US
60/544,835	13 February 2004 (13.02.2004)	US
60/547,933	26 February 2004 (26.02.2004)	US

(71) Applicant *(for all designated States except US)*: **WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH** [US/US]; Nine Cambridge Center, Cambridge, MA 02142-1479 (US).

(72) Inventors; and

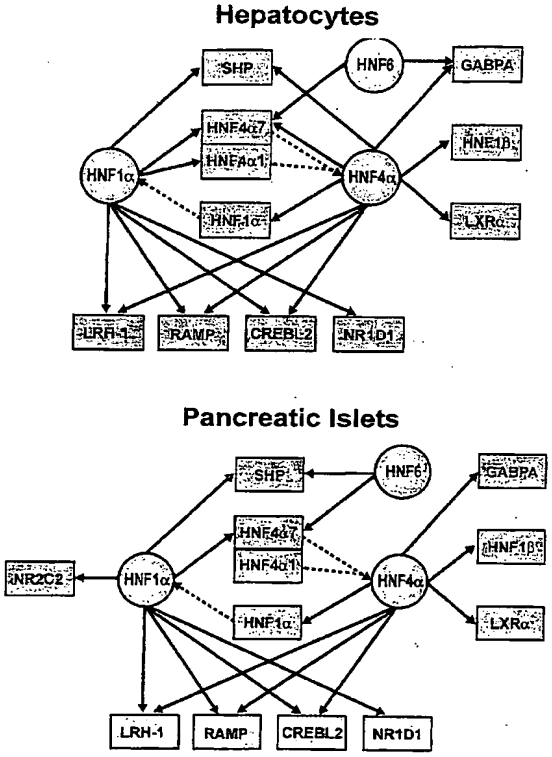
(75) Inventors/Applicants *(for US only)*: **ODOM, Duncan, T.** [US/US]; 48 JFK St. Apt. 20, Cambridge, MA 02138 (US). **YOUNG, Richard, A.** [US/US]; 216 Highland Street, Weston, MA 02193 (US).

(74) Agents: **GRANAHAN, Patricia** et al.; Ropes & Gray LLP, One International Place, Boston, MA 02110-2624 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: TRANSCRIPTIONAL REGULATORS AND METHODS THEREOF



(57) Abstract: The invention relates to transcriptional regulators and related methods thereof. The invention further relates to the identification of genes regulated by transcriptional regulators, to the treatment of diseases associated with abnormal function of a transcriptional regulator and to the modulation of gene expression, including genes expressed in hepatocytes or pancreatic cells, through the modulation of transcriptional regulator activity.

WO 2005/054461 A2

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Transcriptional Regulators and Methods Thereof

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of the filing date of U.S. Application No. 60/525318, filed November 26, 2003, entitled "CONTROL OF PANCREAS AND LIVER GENE EXPRESSION BY HNF TRANSCRIPTION FACTORS", U.S. Application No. 60/542520, filed February 6, 2004, entitled "CONTROL OF PANCREAS AND LIVER GENE EXPRESSION BY HNF TRANSCRIPTION FACTORS", U.S. Application No. 60/544835, filed February 13, 2004, entitled "CONTROL OF PANCREAS AND LIVER GENE EXPRESSION BY HNF TRANSCRIPTION FACTORS", and U.S. Application No. 60/547933, filed February 26, 2004, entitled "TRANSCRIPTIONAL REGULATORS AND METHODS THEREOF". The entire teachings of the referenced applications are incorporated by reference herein.

FUNDING

The invention described herein was supported, in whole or in part, by the U.S. Department of Energy Program for Computational Molecular Biology. The United States government has certain rights in the invention.

BACKGROUND OF THE INVENTION

Gene expression is controlled by transcriptional regulatory proteins, which bind specific DNA sequences and recruit cofactors and the transcription apparatus to promoters (*1-3*). The expression of transcriptional regulators themselves is also regulated by transcriptional regulators, and a single gene may be regulated by multiple transcription factors. As a result of these regulatory networks, or pathways, misregulation of a single transcriptional regulator in a cell can result in the aberrant expression of multiple genes in the network in which the transcriptional regulator is active, leading to disease in the organism.

Current methods of identifying the genes controlled by a transcriptional regulator typically include a comparison of the mRNA levels of candidate target in

cells which express the transcriptional regulator and control cells which either do not express it. Often, this involves overexpressing a recombinant transcriptional regulator in a given cell type and using, as a control cell, one which overexpresses a control recombinant protein or no recombinant protein at all. However, given to the artificial nature of using cell lines and overexpressing transgenes, the results obtained from such approaches may not reflect the *in vivo* regulation by native transcriptional regulators in an organism.

Genome-wide analysis methods have been used recently to determine how tagged transcriptional regulators encoded in *Saccharomyces cerevisae* are associated with the genome in living yeast cells and to model the transcriptional regulatory circuitry of these cells (*4*). These methods have also been used in human tissue culture cells to identify target genes for several transcriptional regulators (*5-7*).

However, the need remains to develop genome-scale analysis methods to determine how transcriptional regulators control the global gene expression programs that characterize specific tissues, and in particular, freshly isolated, primary tissues, in which the transcriptional regulators are likely to maintain their *in vivo* specificities. Furthermore, there is a need to identify the regulatory networks or pathways in which a given transcriptional activator acts, in part, to allow for the identification of therapeutic targets for diseases caused by aberrant function of a transcriptional regulator.

SUMMARY OF THE INVENTION

In one aspect, the invention provides a method of identifying the genes regulated by a transcriptional regulator. One aspect of the invention provides a method of determining which genes from a subset of genes are regulated by a transcriptional regulator in a cell, the method comprising (a) selectively isolating chromatin from a cell which expresses the transcriptional regulator to generate isolated chromatin; (b) selectively isolating chromatin fragments from the isolated chromatin to generate bound chromatin fragments, wherein the bound chromatin fragments are bound by the transcriptional regulator; (c) amplifying both the bound chromatin fragments to generate amplified chromatin fragments and the isolated chromatin to generate

amplified control chromatin; (d) hybridizing the amplified control chromatin and the amplified chromatin fragments to a DNA microarray, wherein the DNA microarray comprises (1) at least 10,000 experimental spots, each experimental spot comprising an experimental DNA, each experimental DNA comprising a promoter region from a gene in the subset; and (2) at least 100 control spots, each control spot comprising a control DNA, each control DNA comprising a non-promoter region; and (e) determining and comparing a hybridization signal at each of the spots on the microarray between those generated by (1) the amplified control chromatin; and (2) the amplified chromatin fragments; wherein a gene in the subset is said to be regulated by the transcriptional regulator in the cell if a spot comprising a promoter region of said gene displays a higher level of hybridization by the amplified chromatin fragments than by the amplified control chromatin.

In another aspect, the invention provides methods of identifying regulatory networks, or pathways, in a cell. The invention provides a method of identifying a transcriptional regulatory network in a cell, the method comprising determining if a transcriptional regulator regulates additional transcriptional regulators in the cell using the method of any of the methods described herein, wherein a transcriptional regulatory network is identified if at least one additional transcriptional regulator is regulated by the transcriptional regulator.

The invention also provides a method of identifying a transcriptional regulatory network in a cell, the method comprising determining if a transcriptional regulator regulates (i) its own promoter; or (ii) a promoter from a plurality of transcriptional regulators; using any of the methods described herein, wherein the experimental DNA comprises (a) a promoter from the transcriptional regulator; and (b) promoters from the plurality of transcriptional regulators;
wherein a transcriptional regulatory network is identified if the transcriptional regulator regulates itself or if it regulates at least one of the plurality of transcriptional regulators.

The invention further provides a method of identifying transcriptional regulatory networks in a cell, the method comprising (a) determining, by repeating a

method of identifying the targets of transcriptional regulator for each of a plurality of transcriptional regulators, the genes in a subset which are regulated by each of the plurality of transcriptional regulators, wherein the experimental DNA comprises promoter regions for each of the plurality of transcriptional regulators; (b) determining if any one of the plurality of transcriptional regulators are regulated by at least one of the plurality of transcriptional regulators; wherein a transcriptional regulatory network is identified if any one of the plurality of transcriptional regulators is regulated by at least one of the plurality of transcriptional regulators.

The invention also provides a DNA microarray for determining promoter occupancy in a human cell, the microarray comprising (1) at least 10,000 experimental spots, each experimental spot comprising an experimental DNA, each experimental DNA comprising a promoter region from a human gene in the subset; and (2) at least 100 control spots, each control spot comprising a control DNA, each control DNA comprising a non-promoter region; wherein at least 75% of the promoter regions comprise from at least 700bp upstream to at least 200 bp downstream of the transcriptional start site.

Another aspect of the invention provides a method of estimating if a transcriptional regulator is a global transcriptional regulator, the method comprising (a) selectively isolating chromatin from a tissue; (b) identifying promoter regions from the chromatin which are bound by a candidate global transcriptional regulator; (c) identifying promoter regions from the chromatin which are bound by a member of the basal transcriptional machinery; and (d) comparing the promoter regions identified in steps (b) and (c) to determine the ratio between (i) the number of promoter regions bound by both the candidate global transcriptional regulator and the member of the basal transcriptional machinery; and (ii) the number of promoter regions bound by the member of the basal transcriptional machinery, wherein a transcriptional regulator is a global transcriptional regulator when the ratio is greater than 0.2.

The invention further provides methods of identifying targets for therapeutics. In one aspect, the invention provides a method of identifying at least one target gene for

the development of a therapeutic to treat or prevent a disorder in a subject, wherein at least one form of the disorder is caused by an altered activity in a transcriptional regulator or in a suspected transcriptional regulator, the method comprising (a) identifying the genes regulated by the transcriptional regulator in a cell; (b) determining if the transcriptional regulator is a broad-acting transcriptional regulator or a narrow-acting transcriptional regulator, wherein if the transcriptional regulator is a broad acting transcriptional regulator then the transcriptional regulator is a target gene for the development of a therapeutic, and wherein if the transcriptional regulator is a narrow acting transcriptional regulator then (i) determining if at least one gene regulated by the transcriptional regulator is likely causative in the disorder, wherein a gene that is likely causative in the disorder is a target gene for the development of a therapeutic; and (ii) reiterating steps (a) and (b) for at least one gene that is regulated by the transcriptional regulator in the cell and that either (1) encodes a transcriptional regulator or (2) is suspected to encode a transcriptional regulator, with the modification that the transcriptional regulator of steps (a) and (b) is said gene, thereby identifying at least one target gene for the development of a therapeutic to treat or prevent a disorder in the subject.

The invention also provides methods of treating or preventing disease. In one aspect, the invention provides a method of treating or preventing type II diabetes in a subject, comprising administering to the subject a therapeutically effective amount of an agent that increases the global transcriptional activity of HNF4alpha.

In another aspect, the invention provides a method of treating or preventing a disorder associated with low transcriptional activity of HNF4alpha in a subject, comprising administering to the subject a therapeutically effective amount of an agent that increases the global transcriptional activity of HNF4alpha. A related aspect provides a method of treating or preventing a disorder associated with high transcriptional activity of HNF4alpha in a subject, comprising administering to the subject a therapeutically effective amount of an agent that decreases the global transcriptional activity of HNF4alpha.

The invention also provides a method of increasing the global transcriptional activity in a liver or a pancreatic cell comprising contacting the cell with an agent which increases the global transcriptional activity of HNF4alpha. A related aspect provides a method of decreasing the global transcriptional activity in a liver or a pancreatic cell comprising contacting the cell with an agent which decreases the global transcriptional activity of HNF4alpha.

One aspect of the invention provides methods of regulating the expression level of genes. On aspect provides a method of regulating the expression level of any one of the genes in Figure 13 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF1alpha. A related aspect provides a method of regulating the expression level of any one of the genes in Figure 14 in a pancreatic cell, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF1alpha.

Another aspect of the invention provides a method of regulating the expression level of any one of the genes in Figure 16 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF6. A related aspect provides a method of regulating the expression level of any one of the genes in Figure 17 in a pancreatic cell, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF6.

Yet another aspect of the invention provides a method of regulating the expression level of any one of the genes in Figure 18 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF4alpha. A related aspect provides a method of regulating the expression level of any one of the genes in Figure 19 in a pancreatic cell, the method comprising contacting the cell with an agent which regulated the transcriptional activity of HNF4alpha.

The invention also provides methods for identifying transcriptionally active genes that are regulated by a transcriptional regulator in a cell. In one aspect, the

invention provides a method of identifying transcriptionally active genes that are regulated by a transcriptional regulator in a cell, the method comprising (a) selectively isolating chromatin from a tissue; (b) identifying promoter regions from the chromatin that are bound by the transcriptional regulator; (c) identifying promoter regions from the chromatin that are bound by a member of the basal transcriptional machinery; and (d) comparing the promoter regions identified in steps (b) and (c) to determine overlapping genes, wherein the overlapping genes are transcriptionally active genes regulated by the transcriptional regulator.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1C show genome-scale location analysis of HNF regulators in human tissues. (A) Hepatocytes and pancreatic islets were obtained from tissue distribution programs. These cells were treated with formaldehyde to covalently link transcription factors to DNA sites of interaction. Cells were harvested, and chromatin in cell lysates was sheared by sonication. The regulator-DNA complexes were enriched by chromatin immunoprecipitation with specific antibodies, the crosslinks were reversed, and enriched DNA fragments and control genomic DNA fragments were amplified using ligation-mediated PCR. The amplified DNA preparations, labeled with distinct fluorophores, were mixed and hybridized onto a promoter array. (B) Venn diagram showing the overlap of HNF1α, HNF6, and HNF4α bound promoters in hepatocytes (top) and pancreatic islets (bottom). (C) The collection of genes occupied by RNA polymerase II in hepatocytes is displayed as a circle, with the genes bound by HNF1α, HNF6, and HNF4α outlined collectively as a fraction of the chart. The relative contributions of HNF1α, HNF6, and HNF4α are shown as framing arcs.

Figures 2A-2B show transcriptional regulatory networks and motifs. (A) HNF1α, HNF6, and HNF4α are at the center of tissue-specific transcriptional regulatory networks. In these examples selected for illustration, regulatory proteins and their gene targets are represented as circles and boxes, respectively. Solid arrows indicate protein-DNA interactions, and genes encoding regulators are linked to their protein products by dashed lines. The HNF4a7 promoter, also known as the P2 promoter (*24, 25*), was recently implicated as a major human diabetes susceptibility locus (see text). (B)

Examples of regulatory network motifs in hepatocytes. For instance, in the multi-component loop, HNF1α protein binds to the promoter of the HNF4α gene, and the HNF4α protein binds to the promoter of the HNF1α gene. These network motifs were uncovered by searching binding data with various algorithms; for details on the algorithms used and a full list of motifs found, see (*20*).

Figure 3 shows one embodiment of a strategy for the identification of at least one target gene of a master regulator for the development of a therapeutic to treat or prevent a disorder.

Figure 4 shows a Venn diagram showing the overlap of two single, independent ChIP experiments using hepatocytes with anti-HNF4α antibodies sc-6556 and sc-8987.

Figure 5 shows a Western blot of HNF4α in HepG2 cells using 50 μg of cell lysate protein with Ab sc-6556. The lower running band is approximately 50 kDa, which is the canonical molecular weight for HNF4a, and the higher running band is the appropriate location for HNF4α dimer. A very similar gel showing HNF4a antibody specificity for sc-6556 is available at the Santa Cruz website (www.scbt.com).

Figures 6A-6D show scatterplots of attempted chromatin immunoprecipitations performed with the anti-HNF4a antibody sc-6556 using Jurkat (T-lymphocyte derived, 6A), BJ-T (foreskin fibroblast derived, 6B), and U937 (histocyte derived, 6C) cells. To demonstrate the noise inherent in the array analysis, applicants show a scatterplot of a sample of input DNA, split, labeled with the two fluorophores, and hybridized to an array (6D). Identical control experiments performed using the anti-HNF1a antibody sc-6547 afforded essentially identical results.

Figure 7 shows a scatterplot of a chromatin immunoprecipitation performed with pre-immune commercial rabbit serum using hepatocytes (left). Goat pre-immune serum and two rabbit sera from different individuals gave a similar scatterplot. For comparison, applicants show the scatterplot for an equivalent ChIP with the anti-HNF4a antibody sc-6556 using hepatocytes (right).

Figure 8 shows a Venn diagram showing the overlap of the sets of promoters bound by HNF4α and RNA Pol II in hepatocytes and pancreatic islets.

Figure 9 shows a composite gel of gene-specific chromatin immunoprecipitation reactions using anti-HNF4α antibody sc-6556 with crosslinked human hepatocytes.

Figure 10 shows composite gel of gene-specific chromatin immunoprecipitation reactions using anti-HNF1α antibody sc-6547 with crosslinked human hepatocytes.

Figure 11 shows a partial list of proximal promoters occupied by of HNF1a in human hepatocytes and pancreatic islets. These genes were assigned to functional categories using the program ProtoGo; genes not in this automated GO ontology database were assigned using Locuslink information. Four genes are shown for each tissue/category combination; for some combinations, fewer than 4 promoters qualified as targets. Hypothetical and functionally uncharacterized genes are not shown. A complete list of targets is available in Figures 13 and 14.

Figure 12 shows Occupancy of BJ-T and tissue-specific promoter sets by HNF factors. (*) Indicates that comparisons between BJ-T and primary tissues used only a subset of Hu13K array promoters, as RNA Pol II was profiled in BJ-T cells using a smaller, prototype array. The denominator in the above fractions represents the number of targets the HNF factor of interest occupied in the set of RNA Pol II occupied promoters that are either BJ-T specific or primary tissue specific.

Figure 13 shows HNF1α bound promoters in hepatocytes

Figure 14 shows HNF1α bound promoters in pancreatic islets.

Figures 15A-15D show genes previously suggested to be regulated by HNF1a and HNF4a. 'Direct' binding is in vivo ChIP and in vivo footprinting, 'in vitro' binding is primarily gel mobility retardation assays and in vitro footprinting, and 'indirect' is

primarily transient transfections. 'Sequence-based' uses a number of different criteria to qualify binding. Note that some duplicate reports are omitted, as are a handful of recent large-scale screens, (e.g. Tronche 1997, Shih 2001, etc.).

Figure 16 shows HNF6 bound promoters in hepatocytes.

Figure 17 shows HNF6 bound promoters in pancreatic islets.

Figure 18A-18C show HNF4α bound promoters in hepatocytes.

Figures 19A-19C show HNF4α bound promoters in pancreatic islets.

Figures 20A-20B show the feed forward regulatory motifs in hepatocytes . The regulatory modules here were derived as described in exemplification. Feed forwards only involving HNF1ą and HNF4ą are also multi-input motifs, as they bind each other's promoters in a multicomponent loop.

Figures 21A-21B show multi-input motifs in hepatocytes. The regulatory modules here were derived as described in the exemplification. MIMs for the HNF6/HNF4ą and HNF1a/HNF4a are listed in Figure 20 as feedforward motifs.

Figures 22A-22B show the feed forward regulatory motifs in pancreatic islets . The regulatory modules here were derived as described in Supporting Online Material. Feed forwards only involving HNF1a and HNF4a are also multiinput motifs, as they bind each other's promoters in a multicomponent loop.

Figures 23A-23B show multi-Input motifs in pancreatic islets The regulatory modules here were derived as described in Supporting Online Material. MIMs for the HNF6/HNF4ą and HNF1a/HNF4ą are listed in Figure 22 as feedforward.

Figures 24A-24B show transcriptional regulators occupied by HNF1ą and HNF4a. Network of DNA regulators downstream of HNF1a and HNF4a in hepatocytes and

islets. Target genes that are among the Gene Ontology "DNA-regulators" category were compiled, and are listed according to functional subcategory.

DETAILED DESCRIPTION OF THE INVENTION

I. Overview

In certain aspects, the invention provides methods related to transcriptional regulators. Some aspects of the invention provide methods for the identification of genes whose transcription is regulated by a specific transcriptional regulator in a cell. Some of these methods comprise determining the promoter occupancy of the transcriptional regulator using a combination of chromatin immunoprecipitation and/or DNA microarray analysis of the promoter regions that are physically associated with the transcriptional regulator in the cell. In some embodiments of the methods described herein, the DNA microarray comprises both experimental spots containing promoter DNA, and control spots containing non-promoter DNA. The methods described herein may be applied to any cell type, including transplant grade primary human tissue. Furthermore, the method described herein can be used to compare the function of transcriptional regulators across cell types, or across two populations, such as healthy and disease-afflicted subjects.

In a related aspect, the invention provides methods of identifying regulatory networks, or pathways. Some methods comprise identifying the transcriptional regulators which are regulated by a given transcriptional regulator, and optionally, determining the genes that are regulated by those transcriptional regulators. Pathways that may be identified using the methods described herein include autoregulatory, multicomponent, feed-forward, and multi-components loops, as well as regulatory chains.

The invention also provides methods of determining if a transcriptional regulator is a global transcriptional regulator. In some aspects, such methods comprise determining the promoter occupancy of both a transcriptional regulator and a member of the basal transcriptional machinery. Comparison of the promoter occupancy by the transcriptional regulator and by the member of the basal transcriptional machinery

allows the identification of transcriptionally active promoters that are bound and regulated by the transcription regulator. Other methods further comprise extrapolating from the set of promoters that were examined to the total number of promoters in the genome to determine the approximate number of transcriptionally active promoters in a cell that are under the control of a specific transcriptional factor or to determine if the transcriptional regulator is a global transcriptional regulator.

Other aspects of the invention provide methods of identifying therapeutic targets to treat disease. One specific aspect of the invention relates to identifying at least one target gene for the development of a therapeutic agent to treat or prevent a disorder in a subject, preferably a disorder in which at least one form of the disorder is caused by an altered activity in a transcriptional regulator or in a gene suspected to encode a transcriptional regulator. Some of the methods provided herein to identify therapeutic targets comprise determining if a transcriptional regulator implicated in the disease is a broad-acting or a narrow-acting transcriptional regulator, such as by identifying at least a subset of the genes that it regulates in a cell, wherein broad-acting transcriptional regulators are targets for therapeutic agents. If the transcriptional regulator is narrow-acting, then the genes that it regulates may be examined further to determine if any are broad-acting transcriptional regulators (for those genes encoding transcriptional regulators) or if any of the genes are causative to the disease state *i.e.* they regulate a pathway or network that is impaired in the disease state.

The invention further provides methods for the treatment of disease. Some aspects of the invention provide methods of treating metabolic disorders, such as type II diabetes. Specific aspects of the invention provide methods of treating or preventing type II diabetes in a subject by administering to the subject a therapeutically effective amount of an agent that increases the global transcriptional activity of HNF4α. Furthermore, the invention provides methods for modulating the expression level of genes. Such methods are based, in part, on the finding by Applicants of genes which are transcriptionally regulated by HNF1α, HNF4α or HNF6 in hepatocytes and pancreatic cells. In a related aspect, the invention provides methods of modulating and expression level of, and alleviating a disease state associated with the abnormal

expression of, the genes in Figures 13-19 by modulating the transcriptional activity or expression of HNF1α, HNF4α or HNF6. In specific embodiments, the expression of the genes is modulated in hepatocytes, pancreatic cells, or both.

II. Definitions

For convenience, certain terms employed in the specification, examples, and appended claims, are collected here. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

The term "including" is used herein to mean, and is used interchangeably with, the phrase "including but not limited" to.

The term "or" is used herein to mean, and is used interchangeably with, the term "and/or," unless context clearly indicates otherwise.

The term "such as" is used herein to mean, and is used interchangeably, with the phrase "such as but not limited to".

A "patient" or "subject" to be treated by the method of the invention can mean either a human or non-human animal, preferably a mammal.

The terms "alpha" and "α" are used interchangeably, as are the terms "beta" and "β".

The term "encoding" comprises an RNA product resulting from transcription of a DNA molecule, a protein resulting from the translation of an RNA molecule, or a protein resulting from the transcription of a DNA molecule and the subsequent

translation of the RNA product.

A "promoter" is a nucleic acid sequence that directs transcription of a nucleic acid. A promoter includes nucleic acid sequences near the start site of transcription, e.g., a TATA box, see, e.g., Butler and Kadonaga (2002) Genes Dev. 16:2583-2592; Georgel (2002) Biochem. Cell Biol. 80:295-300. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs on either side from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions, while an "inducible", promoter is a promoter is active or activated under, e.g., specific environmental or developmental conditions.

The term "expression" is used herein to mean the process by which a polypeptide is produced from DNA. The process involves the transcription of the gene into mRNA and the translation of this mRNA into a polypeptide. Depending on the context in which used, "expression" may refer to the production of RNA, protein or both.

The term "recombinant" is used herein to mean any nucleic acid comprising sequences which are not adjacent in nature. A recombinant nucleic acid may be generated *in vitro*, for example by using the methods of molecular biology, or *in vivo*, for example by insertion of a nucleic acid at a novel chromosomal location by homologous or non-homologous recombination.

The term "transcriptional regulator" refers to a biochemical element that acts to prevent or inhibit the transcription of a promoter-driven DNA sequence under certain environmental conditions (e.g., a repressor or nuclear inhibitory protein), or to permit or stimulate the transcription of the promoter-driven DNA sequence under certain environmental conditions (e.g., an inducer or an enhancer).

The term "microarray" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of

membrane, filter, chip, glass slide, or any other suitable solid support.

The terms "disorders" and "diseases" are used inclusively and refer to any deviation from the normal structure or function of any part, organ or system of the body (or any combination thereof). A specific disease is manifested by characteristic symptoms and signs, including biological, chemical and physical changes, and is often associated with a variety of other factors including, but not limited to, demographic, environmental, employment, genetic and medically historical factors. Certain characteristic signs, symptoms, and related factors can be quantitated through a variety of methods to yield important diagnostic information.

The terms "level of expression of a gene in a cell" or "gene expression level" refer to the level of mRNA, as well as pre-mRNA nascent transcript(s), transcript processing intermediates, mature mRNA(s) and degradation products, encoded by the gene in the cell.

The term "modulation" refers to upregulation (i.e., activation or stimulation), downregulation (i.e., inhibition or suppression) of a response, or the two in combination or apart. A "modulator" is a compound or molecule that modulates, and may be, e.g., an agonist, antagonist, activator, stimulator, suppressor, or inhibitor.

The term "agonist" refers to an agent that mimics or up-regulates (e.g., potentiates or supplements) the bioactivity of a protein, e.g., polypeptide X. An agonist may be a wild-type protein or derivative thereof having at least one bioactivity of the wild-type protein. An agonist may also be a compound that upregulates expression of a gene or which increases at least one bioactivity of a protein. An agonist may also be a compound which increases the interaction of a polypeptide with another molecule, e.g., a target peptide or nucleic acid.

The term "antagonist" refers to an agent that downregulates (e.g., suppresses or inhibits) at least one bioactivity of a protein. An antagonist may be a compound which inhibits or decreases the interaction between a protein and another molecule, e.g., a

target peptide or enzyme substrate. An antagonist may also be a compound that downregulates expression of a gene or which reduces the amount of expressed protein present.

The term "prophylactic" or "therapeutic" treatment refers to administration to the subject of one or more of the subject compositions. If it is administered prior to clinical manifestation of the unwanted condition (e.g., disease or other unwanted state of the host animal) then the treatment is prophylactic, i.e., it protects the host against developing the unwanted condition, whereas if administered after manifestation of the unwanted condition, the treatment is therapeutic (i.e., it is intended to diminish, ameliorate or maintain the existing unwanted condition or side effects therefrom).

The term "therapeutic effect" refers to a local or systemic effect in animals, particularly mammals, and more particularly humans caused by a pharmacologically active substance. The term thus means any substance intended for use in the diagnosis, cure, mitigation, treatment or prevention of disease or in the enhancement of desirable physical or mental development and conditions in an animal or human. The phrase "therapeutically-effective amount" means that amount of such a substance that produces some desired local or systemic effect at a reasonable benefit/risk ratio applicable to any treatment. In certain embodiments, a therapeutically-effective amount of a compound will depend on its therapeutic index, solubility, and the like. For example, certain compounds discovered by the methods of the present invention may be administered in a sufficient amount to produce a reasonable benefit/risk ratio applicable to such treatment.

A probe that is "labeled" is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, isotopic, or chemical means. For example, useful labels include ^{32}P, ^{33}P, ^{35}S, ^{14}C, ^{3}H, ^{125}I, stable isotopes, fluorescent dyes and fluorettes (Rozinov and Nolan (1998) Chem. Biol 5:713-728; Molecular Probes, Inc. (2003) Catalogue, Molecular Probes, Eugene Oreg.), electron-dense reagents, enzymes and/or substrates, e.g., as used in enzyme-linked immunoassays as with those using alkaline phosphatase or horse radish peroxidase. The

label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected. "Radiolabeled" refers to a compound to which a radioisotope has been attached through covalent or non-covalent means. A "fluorophore" is a compound or moiety that absorbs radiant energy of one wavelength and emits radiant energy of a second, longer wavelength.

A "labeled nucleic acid probe or oligonucleotide" is one that is bound, either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the probe can be detected by detecting the presence of the label bound to the probe. The probes are preferably directly labeled as with isotopes, chromophores, fluorophores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex or avidin complex can later bind.

A "nucleic acid probe" is a nucleic acid capable of binding to a target nucleic acid of complementary sequence, usually through complementary base pairing, e.g., through hydrogen bond formation. A probe may include natural, e.g., A, G, C, or T, or modified bases, e.g., 7-deazaguanosine, inosine, etc. The bases in a probe can be joined by a linkage other than a phosphodiester bond. Probes can be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions.

"Small molecule" is defined as a molecule with a molecular weight that is less than 10 kD, typically less than 2 kD, and preferably less than 1 KD. Small molecules include, but are not limited to, inorganic molecules, organic molecules, organic molecules containing an inorganic component, molecules comprising a radioactive atom, synthetic molecules, peptide mimetics; and antibody mimetics. As a therapeutic, a small molecule may be more permeable to cells, less susceptible to degradation, and less apt to elicit an immune response than large molecules. Small molecule toxins are

described, see, e.g., U.S. Pat. No. 6,326,482 issued to Stewart, et al.
A small molecule refers to a composition, which has a molecular weight of less than about 1000 kDa.

III. Identification of Transcriptional Targets and Transcriptional Networks

One aspect of the invention provides a method of determining which genes from a subset of genes are regulated by a transcriptional regulator in a cell, the method comprising (a) selectively isolating chromatin from a cell which expresses the transcriptional regulator to generate isolated chromatin; (b) selectively isolating chromatin fragments from the isolated chromatin to generate bound chromatin fragments, wherein the bound chromatin fragments are bound by the transcriptional regulator; (c) amplifying both the bound chromatin fragments to generate amplified chromatin fragments and the isolated chromatin to generate amplified control chromatin; (d) hybridizing the amplified control chromatin and the amplified chromatin fragments to a DNA microarray, wherein the DNA microarray comprises (1) at least 10,000 experimental spots, each experimental spot comprising an experimental DNA, each experimental DNA comprising a promoter region from a gene in the subset; and (2) at least 100 control spots, each control spot comprising a control DNA, each control DNA comprising a non-promoter region; and (e) determining and comparing a hybridization signal at each of the spots on the microarray between those generated by (1) the amplified control chromatin; and (2) the amplified chromatin fragments; wherein a gene in the subset is said to be regulated by the transcriptional regulator in the cell if a spot comprising a promoter region of said gene displays a higher level of hybridization by the amplified chromatin fragments than by the amplified control chromatin.

Methods of isolating chromatin, and in particular chromatin fragments that are bound by a transcriptional regulator, may be carried out by any method known to one skilled in the art, including by cross-linking the transcriptional regulator to chromatin, fragmenting the chromatin, and immunoprecipitating the transcriptional regulators.

In a preferred embodiment, the chromatin fragments bound by the

transcriptional regulator are isolated using chromatin immunoprecipitation (ChIP). Briefly, this technique involves the use of a specific antibody to immunoprecipitate chromatin complexes comprising the corresponding antigen *i.e.* the transcriptional regulator, and examination of the nucleotide sequences present in the immunoprecipitate. Immunoprecipitation of a particular sequence by the antibody is indicative of interaction of the antigen with that sequence. See, for example, O'Neill et al. in *Methods in Enzymology,* Vol. 274, Academic Press, San Diego, 1999, pp. 189-197; Kuo et al. (1999) *Method* 19:425-433; and Ausubel et al., supra, Chapter 21.

In one embodiment, the chromatin immunoprecipitation technique is applied as follows. Cells which express the transcriptional regulator of interest, such as a native transcriptional regulator or a recombinant transcriptional regulator, are treated with an agent that crosslinks the transcriptional regulator to chromatin if that transcriptional regulator is stably bound to it. In one embodiment of the methods described herein, the crosslinking is formaldehyde crosslinking (Solomon, M.J. and Varshavsky, A., Proc. Natl. Sci. USA 82:6470-6474; Orlando, V., TIBS, 25:99-104). UV light may also be used (Pashev et al. *Trends Biochem Sci.* 1991;16(9):323-6; Zhang L et al. *Biochem Biophys Res Commun.* 2004;322(3):705-11).

Subsequent to crosslinking, cellular nucleic acid is isolated, sheared such as by sonication and incubated in the presence of an antibody directed against the transcriptional regulator. Antibody-antigen complexes are precipitated, crosslinks are reversed (for example, formaldehyde-induced DNA-protein crosslinks can be reversed by heating) so that the sequence content of the immunoprecipitated DNA is tested for the presence of a specific sequence, for example, promoter regions. The antibody may bind directly to an epitope on the transcriptional regulator or it may bind to a tag on the regulator, such as a myc tag when used with an anti-Myc antibody (Santa Cruz Biotechnology, sc-764).

In yet another embodiment, a non-antibody agent with affinity for the transcriptional regulator or for a tag used to it is used in place of the antibody. For example, if the transcriptional regulator comprises an affinity tag, such as a six-

histidine tag, complexes may be isolated by affinity chromatography to nickel-containing sepharose. Additional variations on ChIP methods within the scope of the invention may be found in Kurdistani et al. Methods. 2003 31(1):90-5; O'Neill et al. Methods. 2003, 31(1):76-82; Spencer et al., Methods. 2003;31(1):67-75; and Orlando et al. Methods 11: 205-214 (1997).

In an alternate embodiment of the methods described herein for identifying genes regulated by a transcriptional regulator, amplified chromatin fragments from a control immunoprecipitation reaction are used in place of the isolated chromatin as a control. For example, an antibody that does not react with the transcription factor being tested may be used in a chromatin IP procedure to isolate control chromatin, which can then be compared to the chromatin isolated using an antibody that does react with the transcriptional regulator. In preferred embodiments, the antibody that does not react with the transcription factor being tested also does not react with other transcriptional regulators or DNA binding proteins.

In one embodiment, the amplified control chromatin and the amplified chromatin fragments are generated from their corresponding template DNA using ligation-mediated polymerase chain reaction (LM-PCR) (e.g., see Current Protocols in Molecular Biology, Ausubel, F. M. et al., eds. 1991, and U.S. Application No. 2003/0143599, the teachings of which are incorporated herein by reference) in their entirety. In specific embodiments, LM-PCR comprises fluorescently labeling amplified DNA by including fluorescently-tagged nucleotides in the LM-PCR reaction. Additional variations for manipulating and examining chromatin using microarrays have described in U.S. Patent Nos. 6,410,243, the teachings of which are incorporated herein by reference.

In one embodiment, the labelled or unlabeled probes are hybridized to DNA microarray, such as is described in U.S. Patent No. 6,410,243. Microarrays, also called "biochips" or "arrays" are miniaturized devices typically with dimensions in the micrometer to millimeter range for performing chemical and biochemical reactions and are particularly suited for embodiments of the invention. Arrays may be constructed via

microelectronic and/or microfabrication using essentially any and all techniques known and available in the semiconductor industry and/or in the biochemistry industry, provided only that such techniques are amenable to and compatible with the deposition and screening of polynucleotide sequences. Microarrays are particularly desirable for their virtues of high sample throughput and low cost for generating profiles and other data. Additional variations for manipulating and examining chromatin using microarrays have described in U.S. Patent Nos. 6,410,243, the teachings of which are incorporated herein by reference.

In one embodiment of the methods described, amplified control chromatin and the amplified chromatin fragments are hybridized to a DNA microarray that includes experimental spots that represent all or a subset (e.g., a chromosome or chromosomes) of the genome. The fluorescent intensity of each experimental spot on the microarray from the amplified chromatin fragments relative to the amplified control chromatin indicates whether the protein of interest is bound to the DNA region located at that particular spot. Hence, the methods described herein allow the detection of protein-DNA interactions across an entire genome.

In some embodiments of the methods described herein, the promoter region of a gene comprises from at least 700bp upstream to at least 200 bp downstream of the transcriptional start site of the gene. In some embodiments, the promoter region comprises at least about 30, 40, 50, or 60 nucleotides in length. In specific embodiments, the promoter region of a gene as found on the spots of the microarray comprises a sequence of at least 30 nucleotides whose sequence is identical to a region stretching from 3 kb upstream to 1 kb downstream of the transcriptional start site of said gene. In some embodiments, the DNA microarray includes control spots of non-promoter DNA. In specific embodiment, the non-promoter region comprises an open reading frame. In preferred embodiments, the non-promoter regions comprise genomic regions which are not bound by transcriptional regulators, and preferably which are not bound by the transcriptional regulator being tested. In some embodiments, not all the experimental spots or the control spots comprise experimental DNA or control DNA, respectively. Furthermore, in some specific embodiments some spots comprise control

DNA which comprises promoter DNA. One skilled in the art may determine the number of experimental or control spots for a given application.

In some embodiments of the methods described herein, the level of hybridization of the amplified chromatin fragments to each experimental spot is normalized by the level of hybridization of the amplified chromatin fragments to the control spots. In specific embodiments, the normalization is performed by subtracting the mean level of hybridization of the amplified chromatin fragments to the control spots from the level of hybridization of the amplified chromatin fragments at each experimental spot.

Methods of analyzing data from microarrays are well-described in the art, including in DNA Microarrays: A Molecular Cloning Manual, Ed by Bowtel and Sambrook (Cold Spring Harbor Laboratory Press, 2002); Microarrays for an Integrative Genomics by Kohana (MIT Press, 2002); A Biologist's Guide to Analysis of DNA Microarray Data, by Knudsen (Wiley, John & Sons, Incorporated, 2002); and DNA Microarrays: A Practical Approach, Vol. 205 by Schema (Oxford University Press, 1999); and Methods of Microarray Data Analysis II, ed by Lin et al. (Kluwer Academic Publishers, 2002), hereby incorporated by reference in their entirety.

In some embodiments of any of the methods described herein, the transcriptional regulator is native to the cell. By native it is meant that the transcriptional regulator naturally occurs in the cell. In other embodiments, the transcriptional regulator is a recombinant transcriptional regulator. In some embodiments, the transcriptional regulator originates from a species which is different from that of the cell. In some embodiments, the transcriptional regulator is a viral transcriptional regulator. In such embodiments, a cell may be contacted with a virus and chromatin extracted from the infected cell after allowing sufficient time for the viral proteins to be expressed. In some embodiments, recombinant transcriptional regulators have missense mutations, truncations, or inserted sequences or entire domains from other naturally occurring proteins. A tagged recombinant transcriptional regulator may be used in some embodiments the methods of the present invention as

the tag may facilitate the immunoprecipitation of the regulator.

In certain embodiments of the invention, transcriptional regulators comprise specific transcription factors, coactivators, corepressors or complexes thereof. Transcription factors bind to specific cognate DNA elements such as promoters, enhancers and silencer elements, and are responsible for regulating gene expression. Transcription factors may be activators of transcription, repressors of transcription or both, depending on the cellular context. Transcription factors may belong to any class or type of known or identified transcription factor. Examples of known families or structurally-related transcription factors include helix-loop-helix, leucine zipper, zinc finger, ring finger, and hormone receptors. Transcription factors may also be selected based upon their known association with a disease or the regulation of one or more genes. For example, transcription factors such as c-myc, Rel/Nf-kB, neuroD, c-fos, c-jun, and E2F may be targeted. Antibodies directed to any transcriptional coactivator or corepressor may also be used according to the invention. Examples of specific coactivators include CBP, CTIIA, and SRA, while specific examples of corepressors include the mSin3 proteins, MITR, and LEUNIG. Furthermore, the genes regulated by proteins associated with transcriptional complexes, such as the histone acetylases (HATs) and histone deacetylases (HDACs), may also de determined using the methods described herein.

In one embodiment of the methods described herein, the cell is a primary cell. Primary cells are directly isolated from an organism and have undergone minimum passaging *in vitro*, and thus maintain most of the phenotypic characteristics of cells in the organism. In a specific embodiment, the primary cells are primary cells that have doubled less than 10 times *ex vivo*. In some embodiments, the cell is derived from transplant grade tissue or freshly isolated tissue. The cell type used in the assays described herein may be any cell type. The cell may be eukaryotic or prokaryotic, from a metazoan or from a single-celled organism such as yeast. In some preferred embodiments the cell is a mammalian cell, such as a cell from a rodent, a primate or a human. The cell may be a wild-type cell or a cell that has been genetically modified by recombinant means or by exposure to mutagens. The cell may be a transformed cell or

an immortalized cell. In some embodiments, the cell is from an organism afflicted by a disease. In some embodiments, the cell comprises a genetic mutation that results in disease, such as in a hyperplastic condition.

In some embodiments, the cell is derived from transplant-grade tissue or freshly isolated tissue. In some embodiments, the cell is derived from a tissue biopsy, such as from a subject afflicted with, or suspected of being afflicted with, a disorder. In another embodiment, the cell is isolated from a bodily fluid or bodily secretion, including serum, plasma, saliva, tears, sweat, semen, amniotic fluid, vaginal secretions, nasal secretions, synovial fluid, spinal fluid, phlegm, bronchoalveolar lavage fluid, blister fluid, pus, stool and intracranial fluid. The cell may be a live cell or a cell that has been preserved, such as by treatment with formalin, B5, Zenker's fixatives, Lugol's solution, Carnoy's Fixative, F13 fixative, or other preservatives, or a cell that has been preserved by freezing.

In some embodiments of the methods described herein, the cell has been treated with an agent, such as compound or a drug, prior to isolation of chromatin. Some preferred agents include those which bind to or regulate the expression of transcriptional regulators. In some embodiments, the genes that are regulated by a given transcriptional regulator are determined both in a cell that is contacted with an agent and in a cell that is not contacted with the agent, or that is contacted with a different amount of the agent. Such methods may be used to identify compounds that alter the types of genes and/or the extent to which a transcriptional regulators controls transcription of those genes. Furthermore, such approaches may be used to screen for agents which alter the activity, specificity or expression of a transcriptional regulator.

In some embodiment of the methods described herein for identifying genes regulated by a transcriptional regulator, a higher level of hybridization by the amplified chromatin fragments than by the amplified control chromatin comprises at least a two-fold higher level of hybridization. The threshold for what constitutes a higher level of hybridization, may be adjusted by one skilled in the art for the particular application. Higher levels of hybridization are expected to yield a smaller target size but with higher

certainty that a given gene above that threshold is regulated by the transcriptional regulator in that cell *in vivo*.

In other embodiments of the methods described herein for identifying genes regulated by a transcriptional regulator, the transcriptional regulator is a basal transcription factor or a component of the basal transcription machinery. In specific embodiments, components of the basal transcription machinery comprise RNA polymerases, including poII, poIII and poIIII, TBP, NTF-1 and Sp1 and any other component of TFIID, including, for example, the TAFs (e.g. TAF250, TAF150, TAF135, TAF95, TAF80, TAF55, TAF31, TAF28, and TAF20), or any other component of a polymerase holoenzyme.

Another aspect of the invention provides a method of identifying transcriptionally active genes that are regulated by a transcriptional regulator in a cell. The method comprises determining what genes are regulated by the transcriptional regulator and determining which ones are transcriptionally active in the cell. In one embodiment, a set of genes which are transcriptionally active is the set of genes whose promoters are bound by an RNA polymerase, such as RNA polymerase II, or by a member of the basal transcription machinery. Alternatively, genes which are transcriptionally active may be identified using other techniques know in the art. For example, mRNA from a cell which expresses the transcriptional regulator can be collected and examined on a DNA microarray which comprises coding sequences in order to determine which genes are being transcribed.

In one embodiment, the invention provides a method of identifying transcriptionally active genes that are regulated by a transcriptional regulator in a cell, the method comprising (a) selectively isolating chromatin from a tissue; (b) identifying promoter regions from the chromatin that are bound by the transcriptional regulator; (c) identifying promoter regions from the chromatin that are bound by a member of the basal transcriptional machinery; and (d) comparing the promoter regions identified in steps (b) and (c) to determine overlapping genes, wherein the overlapping genes are transcriptionally active genes regulated by the transcriptional regulator.

In a related aspect, the invention provides methods to determine if a transcriptional regulator is a global transcription regulator. One method comprises estimating if a transcriptional regulator is a global transcriptional regulator, the method comprising (a) selectively isolating chromatin from a tissue; (b) identifying promoter regions from the chromatin which are bound by a candidate global transcriptional regulator; (c) identifying promoter regions from the chromatin which are bound by a member of the basal transcriptional machinery; and (d) comparing the promoter regions identified in steps (b) and (c) to determine the ratio between (i) the number of promoter regions bound by both the candidate global transcriptional regulator and the member of the basal transcriptional machinery; and (ii) the number of promoter regions bound by the member of the basal transcriptional machinery wherein a transcriptional regulator is a global transcriptional regulator when the ratio is greater than 0.2.

In a preferred embodiment of the methods described above, steps (b) and (c) are performed using a DNA microarray. In a specific embodiment, the DNA microarray comprises (i) at least 10,000 experimental spots, each experimental spot comprising an experimental DNA, each experimental DNA comprising a promoter region from a human gene in the subset; and (ii) at least 100 control spots, each control spot comprising a control DNA, each control DNA comprising a non-promoter region. Any type of microarray or array may be used.

In one embodiment of the methods described above, the member of the transcriptional machinery is an RNA polymerase, such as RNA polymerase II, a TATA-binding protein, or any other component of TFIID, including, for example, the TAFs (e.g. TAF250, TAF150, TAF135, TAF95, TAF80, TAF55, TAF31, TAF28, and TAF20).

Another aspect of the invention provides methods of identifying regulatory networks, or pathways, in a cell. The methods provided by the invention allow the identification of the regulatory motifs, such as those shown in Figure 2B. A regulatory pathway can include, for example, a pathway that controls a cellular function under a

specific condition. A regulatory pathway controls a cellular function by, for example, altering the activity of a system component or the activity of a biochemical, gene expression or other type of pathway. Alterations in activity include, for example, inducing a change in the expression, activity, or physical interactions of a pathway component under a specific condition. Specific examples of regulatory pathways include a pathway that activates a cellular function in response to an environmental stimulus of a biochemical system, such as the inhibition of cell differentiation in response to the presence of a cell growth signal and the activation of galactose import and catalysis in response to the presence of galactose and the absence of repressing sugars. The term "component" when used in reference to a network or pathway is intended to mean a molecular constituent of the biochemical system, network or pathway, such as, for example, a polypeptide, nucleic acid, other macromolecule or other biological molecule.

In one aspect, the invention provides a method of identifying a transcriptional regulatory network in a cell, the method comprising determining if a transcriptional regulator regulates additional transcriptional regulators in the cell, such as by using any of the methods described herein, wherein a transcriptional regulatory network is identified if at least one additional transcriptional regulator is regulated by the transcriptional regulator.:

Another aspect of the invention provides a method of identifying a transcriptional regulatory network in a cell, the method comprising determining if a transcriptional regulator regulates (i) its own promoter; or (ii) a promoter from a plurality of transcriptional regulators; such as by using any of the methods described herein, wherein the experimental DNA comprises (a) a promoter from the transcriptional regulator; and (b) promoters from the plurality of transcriptional regulators; wherein a transcriptional regulatory network is identified if the transcriptional regulator regulates itself or if it regulates at least one of the plurality of transcriptional regulators.

Yet another aspect of the invention provides a method of identifying

transcriptional regulatory networks in a cell, the method comprising (a) determining, by repeating one of the methods described herein for each of a plurality of transcriptional regulators, the genes in a subset which are regulated by each of the plurality of transcriptional regulators, wherein the experimental DNA comprises promoter regions for each of the plurality of transcriptional regulators; (b) determining if any one of the plurality of transcriptional regulators are regulated by at least one of the plurality of transcriptional regulators; wherein a transcriptional regulatory network is identified if any one of the plurality of transcriptional regulators is regulated by at least one of the plurality of transcriptional regulators.

Specific embodiments of the methods for identifying regulatory networks described herein further comprise determining if any of the genes regulated by one of the plurality of transcriptional regulators is also a target of any of the other transcriptional regulators

The invention further provides algorithms for the identification of regulatory motifs, which may be used in conjuction with any of the methods provided herein, such as the methods for identifying the genes regulated by a transcriptional regulator. In a specific embodiment, two data matrices are created. The overall matrix D consists of binary entries Dij, where a 1 indicates binding of regulator j to intergenic region i, a 0 indicates no binding event. The regulator matrix R is a subset of D, containing only the rows corresponding to the intergenic region assigned to each regulator, in the same order as the columns of regulators. The analyses may be performed using Matlab® software. The algorithms to find each motif are described as follows:

Autoregulatory motif: Find each non-zero entry on the diagonal of R.

Feedforward loop: For each master regulator (column of R), find non-zero entries, which correspond to regulators bound. For each master regulator / secondary regulator pair, find all rows in D bound by both regulators.

Multi-component loop: For each regulator (column of R), find the regulators to

which it binds. For each of these, find the regulators it binds. If any of these are the original regulator, you have a multi-component loop of two. For all others, find regulators to which they bind. If any of these are the original, you have a multicomponent loop of three. Repeat to find larger loops.

Single input module: Find the intergenic regions bound by only one regulator. That is, take the subset of rows of D such that the sum of each row is 1. Then for each regulator (column), find non-zero entries. Each set (greater than three intergenic regions) is a SIM.

Multi-input module: Find the intergenic regions bound by more than one regulator. That is, take the subset of rows of D such that the sum of each row is greater than 1. Then, for each row, find any other row bound by the same regulators. The collection of rows bound by the same regulators correspond to a MIM. Once a row is assigned to a MIM, remove it from further analysis.

Regulator chain: For each regulator (column of R), use a recursive algorithm to find chains of all lengths. That is, for each regulator whose promoter is bound by the regulator before it in the chain, find the regulator promoters to which it binds. Repeat until the chain ends. There are three possible ways to end a chain: a regulator that does not bind to the promoter of any other regulator, a regulator that binds to its own promoter, or one that binds to the promoter of another regulator earlier in the chain.

In one preferred embodiment of any of the methods described herein such as the methods for identifying regulatory networks, the experimental DNA in the microarray comprises promoter regions from additional transcriptional regulators or from genes suspected to encode transcriptional regulators. Such microarray enables one skilled in the art to identify the components of a regulatory pathway. For example, starting with one transcriptional regulator, a subset of the genes it regulates are identified using any method, such as those described herein. If one identified gene is itself a second transcriptional regulator or is suspected to encode a transcriptional regulator, then the subset of genes the second transcriptional regulator regulates is identified, and so on.

Furthermore, the subset of genes that the first and second transcriptional regulators regulate can be compared to determine of any genes are found in both subsets. If so, then a feed-forward motif, a unit of a regulatory network, has been identified. Likewise, if the second transcriptional regulator is found to regulate the first one, then a feedback loop has been identified.

4. Development of a Therapeutic to Treat or Prevent Disorders

One aspect of the invention provides methods of identifying targets for the development of therapeutics. One aspect of the invention provides a method of identifying at least one target gene for the development of a therapeutic to treat or prevent a disorder in a subject, wherein at least one form of the disorder is caused by an altered activity in a transcriptional regulator or in a suspected transcriptional regulator, the method comprising (a) identifying the genes regulated by the transcriptional regulator in a cell; (b) determining if the transcriptional regulator is a broad-acting transcriptional regulator or a narrow-acting transcriptional regulator, wherein if the transcriptional regulator is a broad acting transcriptional regulator then the transcriptional regulator is a target gene for the development of a therapeutic, and wherein if the transcriptional regulator is a narrow acting transcriptional regulator then (i) determining if at least one gene regulated by the transcriptional regulator is likely causative in the disorder, wherein a gene that is likely causative in the disorder is a target gene for the development of a therapeutic; and (ii) reiterating steps (a) and (b) for at least one gene that is regulated by the transcriptional regulator in the cell and that either (1) encodes a transcriptional regulator or (2) is suspected to encode a transcriptional regulator, with the modification that the transcriptional regulator of steps (a) and (b) is said gene, thereby identifying at least one target gene for the development of a therapeutic to treat or prevent a disorder in the subject.

In some embodiments of the methods for identifying a target gene for the development of a therapeutic, the genes regulated by the transcriptional regulator in the cell are identified using chromosome-wide location analysis, analysis of mRNA transcripts in a cell that expresses the transcriptional regulator, or by using any of the methods provided herein for the identification of the genes that are regulated by a

transcriptional regulator. Some methods may comprise the use of DNA microarray or DNA arrays, such as those described in Gabrielson et al., Obesity Research, 8(5), 374-384 (2000).

In some embodiments of the methods described herein for identifying a target gene for the development of a therapeutic, the transcriptional regulator is a master regulatory gene. In specific embodiments, the master regulatory gene is SOX1-18, OCT6, PAX3, Myocardin, GATA1-6, TCF1/HNF1A, HNF4A, HNF6, NGN3, C/EBP, FOXA1-3, IPF1, GATA, HNF3, NKX2.1, CDX, FTF/NR5A2, C/EBPbeta, SCL1, SKIN1, or a member of the neurogenin, LK, LMO, SOX, OCT, PAX, GATA or MyoD family of transcription factors.

In some embodiments of the methods described herein, the transcriptional regulator is PAX3, EGR-1, EGR-2, OCT6, a SOX family member, a GATA family member, a PAX family member, an OCT family member, RFX5, WHN, GATA1, VDR, CRX, CBP, MeCP2, AML1, p53, PLZF, PML, Rb, WT1, NR3C2, GCCR, PPARgamma, SIM1, HNF1alpha, HNF1beta, HNF4alpha, PDX1, MAFA, FOXA2, or NEUROD1.

A transcriptional regulator whose altered activity can lead to disease might be expressed in multiple, or all tissues of an organism, such that any of multiple cell types may be used in identifying a therapeutic. In some embodiments of the methods described herein for identifying a target gene for the development of a therapeutic, the cell is derived from a tissue whose function is impaired in the disorder. For example, a pancreatic cell may be used for diabetes, a cardiac muscle cells for myocardial infarction, or neurons for Alzheimer's disease.

In specific embodiments of the methods described herein for identifying a target gene for the development of a therapeutic, the broad acting gene regulates at least about 1%, 2% or more preferably at least about 2.5% of the genes in the cell, and the narrow acting gene regulates less than about 1%, 2% or 2.5% of the genes in the cell.

In specific embodiments of the methods described herein, a gene is suspected to encode a transcriptional regulator if it shares at least about 30%, 40% or 50% amino acid sequence identity within at least the DNA binding domain of a transcriptional regulator. DNA binding domains and methods of performing nucleic acids and polypeptide sequence alignments are well-known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2: 482 (1981); by the homology alignment algorithm of Needleman and Wunsch, *J. Mol Biol.* 48: 443 (1970); by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* 8: 2444 (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, Calif., GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 7 Science Dr., Madison, Wis., USA; the CLUSTAL program is well described by Higgins and Sharp, *Gene,* 73: 237-244, 1988; Higgins and Sharp, CABIOS :11-13, 1989; Corpet, et al., *Nucleic Acids Research,* 16:881-90,1988; Huang, et al., *Computer Applications in the Biosciences* 8:1-7,1992; and Pearson, et al., *Methods in Molecular Biology* 24:7-331,1994.

In some specific embodiments of the methods described herein for identifying a target gene for the development of a therapeutic, the gene regulated by the transcriptional regulator is said to be likely causative of the disorder if a mutation in said gene results in at least one phenotype or symptom associated with the disorder. In another specific embodiment, the gene regulated by the transcriptional regulator is said to be likely causative of the disorder when the gene encodes an enzyme or signaling molecule which functions in a pathway that is impaired in the disorder. For example, if the disease is type II diabetes, a disorder characterized by hyperglycemia, then a gene regulated by the transcriptional regulator which encodes a sugar transporter, an enzyme involved in catalyzing a step of glycolysis or gluconeogenesis, or a gene which regulates insulin production, secretion or signaling is said to be likely causative or the disorder. In another specific embodiment, the gene regulated by the transcriptional regulator is said to be likely causative of the disorder if a mutant allele of the gene is genetically linked to a "susceptibility locus" for at least one form of the disease. A

"susceptibility locus" for a particular disease is a sequence or gene locus implicated in the initiation or progression of the disease. The susceptibility locus can be, for example, a gene or a microsatellite repeat, as identified by a microsatellite marker, or can be identified by a defined single nucleotide polymorphism. Generally, susceptibility genes implicated in specific diseases and their loci can be found in scientific publications, but may also be determined experimentally.

In some embodiments of the methods described herein for identifying a target gene for the development of a therapeutic, the altered activity in the transcriptional regulator comprises at least one of the following: (a) an alteration in the binding affinity of the transcriptional regulator to DNA; (b) an alteration in the ability of the transcriptional regulator to bind to RNA polymerase, to an RNA polymerase holoenzyme, or to a second transcriptional regulator; (c) an alteration in the binding affinity of the transcriptional regulator to a ligand; (d) an alteration in expression level or expression pattern of the transcriptional regulator; or (e) an alteration in an ability of the transcriptional regulator to form homomultimers or heteromultimers.

In some embodiments of the methods described herein, the cell comprises a mutant form of the transcriptional regulator. A preferred mutant form of the transcriptional regulator is one that causes the disease to which the therapeutic is sought. Such embodiments are particularly preferred when a mutant transcriptional regulator which causes at least one form of the disease has an altered target specificity and thus the genes it regulates, or the extent to which it regulates their transcription, is altered when compared to the non-mutant form of the transcriptional regulator. Such embodiments may allow the identification of therapeutic targets which might not have been identified if a wild-type form of the transcriptional regulator had been used. Mutations in the DNA binding domain, for example, may alter the target specificity of a transcriptional regulator by altering its affinity for various DNA binding sequences.

It is well-known to one skilled in the art that mutations in a transcriptional regulator may result in a hypomorphic, hypermorphic or neomorphic phenotype. Mutations may generally reduce the activity of a transcriptional regulator, may

generally increase it activity, or may confer novel properties, such as altering the range of targets or turning an activator into a repressor or vice versa. In any methods described herein, and in particular those for identifying the therapeutics, a cell expressing a transcriptional regulator having any of these changes in activity may be used.

The methods described herein may be applied to any disorder for which a transcriptional regulator has been implicated. Examples of diseases and transcriptional regulators which cause them may be found in the scientific and medical literature by one skilled in the art, including in Medical Genetics, L.V. Jorde et al., Elsevier Science 2003, and Principles of Internal Medicine, 15th edition, ed by Braunwald et al., McGraw-Hill, 2001; American Medical Association Complete Medical Encyclopedia (Random House, Incorporated, 2003); and The Mosby Medical Encyclopedia, ed by Glanze (Plume, 1991). In some embodiments, the disorder is characterized by impaired function of at least one of the following: brain, spinal cord, heart, arteries, esophagus, stomach, small intestine, large intestine, liver, pancreas, lungs, kidney, urinary tract, ovaries, breasts, uterus, testis, penis, colon, prostate, bone, muscle, cartilage, thyroid gland, adrenal gland, pituitary, bone marrow, blood, thymus, spleen, lymph nodes, skin, eye, ear, nose, teeth or tongue.

In some embodiments of the methods described herein for identifying a target gene for the development of a therapeutic, the subject is a mammal. In preferred embodiments, the subject is a human. In some embodiments of the methods described herein for identifying a target gene for the development of a therapeutic, the therapeutic comprises a small molecule drug, an antisense nucleic acid, an antibody, a peptide, a ligand, a fatty acid, a hormone or a metabolite.

Antisense nucleic acids acting by RNAi include oligonucleotides which specifically hybridize (e.g., bind) under cellular conditions with a gene sequence, such as at the cellular mRNA and/or genomic DNA level, so as to inhibit expression of that gene, e.g., by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarily, or, for example, in the case of binding to DNA

duplexes, through specific interactions in the major groove of the double helix. Preferred antisense nucleic acid comprise siRNA, shRNAs, or any other form of double stranded RNA molecule. Antisense nucleic acids may be chemically modified, such as to increase their in vivo stability.

RNAi is a process of sequence-specific post-transcriptional gene repression which can occur in eukaryotic cells. In general, this process involves degradation of an mRNA of a particular sequence induced by double-stranded RNA (dsRNA) that is homologous to that sequence. For example, the expression of a long dsRNA corresponding to the sequence of a particular single-stranded mRNA (ss mRNA) will labilize that message, thereby "interfering" with expression of the corresponding gene. Accordingly, any selected gene may be repressed by introducing a dsRNA which corresponds to all or a substantial part of the mRNA for that gene. It appears that when a long dsRNA is expressed, it is initially processed by a ribonuclease III into shorter dsRNA oligonucleotides of in some instances as few as 21 to 22 base pairs in length. Furthermore, RNAi may be effected by introduction or expression of relatively short homologous dsRNAs. dsRNAs shorter than about 30 bases pairs are preferred to effect gene repression by RNAi (see Hunter et al. (1975) J Biol Chem 250: 409-17; Manche et al. (1992) Mol Cell Biol 12: 5239-48; Minks et al. (1979) J Biol Chem 254: 10180-3; and Elbashir et al. (2001) Nature 411: 494-8).

Antibodies include whole antibodies, e.g., of any isotype (IgG, IgA, IgM, IgE, etc.), and includes fragments thereof which are also specifically reactive with a vertebrate, e.g., mammalian, protein. Antibodies may be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. Thus, the term includes segments of proteolytically-cleaved or recombinantly-prepared portions of an antibody molecule that are capable of selectively reacting with a certain protein. Non-limiting examples of such proteolytic and/or recombinant fragments include Fab, F(ab')2, Fab', Fv, and single chain antibodies (scFv) containing a V[L] and/or V[H] domain joined by a peptide linker. The scFv's may be covalently or non-covalently linked to form antibodies having two or more binding sites. The subject invention includes polyclonal, monoclonal,

humanized, or other purified preparations of antibodies and recombinant antibodies.

Peptidomimetic include compounds containing peptide-like structural elements that is capable of mimicking the biological action (s) of a natural parent polypeptide.

Hormone include any one of a number of biochemical substances that are produced by a certain cell or tissue and that cause a specific biological change or activity to occur in another cell or tissue located elsewhere in the body.

Metabolites includes any substance produced by metabolism or by a metabolic process. "Metabolism", as used herein, refers to the various chemical reactions involved in the transformation of molecules or chemical compounds occurring in tissue and the cells therein.

Ligands include any substance which binds to a receptor protein. A ligand of a transcriptional regulator protein is a substance which binds to the regulator protein, such as estrogen binding to a nuclear hormone receptor. In a preferred embodiment, ligand binding of to a transcriptional regulator occurs with high affinity. The term ligand refers to substances including, but not limited to, a natural ligand, whether isolated and/or purified, synthetic, and/or recombinant, a homolog of a natural ligand (e.g., from another mammal). The term ligand encompasses substances which are inhibitors or promoters of receptor activity, as well as substances which selectively bind receptors, but lack inhibitor or promoter activity.

Some aspects of the invention relate to the diagnosis of disease states. A "transcriptional fingerprint", or listing of the genes, and optionally to what extent, that are regulated by given a transcriptional regulator can be generated from healthy individuals and from those afflicted with a disorder. Comparison of the fingerprints between the two groups may define genes which are specific to one of the two groups, and thus serve as diagnostic for the risk that a patient is at risk, or is afflicted, with the disorder. In one embodiment, the transcriptional fingerprint of HNF4a is used to diagnose type II diabetes. A biopsy of a subject's liver or pancreas may provide the

cells for such analysis.

In specific embodiments, the transcriptional fingerprint disease diagnosis analysis is applied to transcriptional regulators which are causative in a particular disease to diagnose the disease. This approach may be coupled to allelic genotyping of the transcriptional regulator gene in the subject. For example, genotyping of a subject's HNF4a may uncover a novel allele. By using "transcriptional fingerprint" of HNF4a in tissue from that patient, one skilled in the art may determine what effect that mutation has in HNF4a activity and thus diagnose type II diabetes.

5. Methods of Preventing/Treating Disease through Regulation of HNFs

Some aspects of the invention provide methods of treating or preventing disease by regulating transcriptional regulator activity, particularly that of the HNF family member. The invention provides a method of treating or preventing type II diabetes in a subject, comprising administering to the subject a therapeutically effective amount of an agent that increases the global transcriptional activity of HNF4alpha. U.S. Patent No. 5,849,485 describes methods and assays for the isolation of modulators of HNF-4a activity, hereby incorporated by reference.

The invention also provides a method of treating or preventing a disorder associated with low transcriptional activity of HNF4alpha in a subject, comprising administering to the subject a therapeutically effective amount of an agent that increases the global transcriptional activity of HNF4alpha. In a related aspect, the invention provides a method of treating or preventing a disorder associated with high transcriptional activity of HNF4alpha in a subject, comprising administering to the subject a therapeutically effective amount of an agent that decreases the global transcriptional activity of HNF4alpha.

Yet another related aspect of the invention provides a method of increasing the global transcriptional activity in a liver or a pancreatic cell comprising contacting the cell with an agent which increases the global transcriptional activity of HNF4alpha. Similarly, the invention provides a method of decreasing the global transcriptional

activity in a liver or a pancreatic cell comprising contacting the cell with an agent which decreases the global transcriptional activity of HNF4alpha.

Applicants have identified genes that are transcriptionally regulated by HNF-1a, HNF4a and HNF6 in hepatocytes and pancreatic cells. Accordingly, the invention provides methods of regulating the expression level of any of these genes in a cell or in a subject by contacting the cell or administering to the subject and agent which modulates the expression level or transcriptional regulatory activity of HNF transcription factors.

The invention provides a method of regulating the expression level of any one of the genes in Figure 13 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF1alpha. Similarly, the invention also provides a method of regulating the expression level of any one of the genes in Figure 14 in a pancreatic cell, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF1alpha.

The invention also provides a method of regulating the expression level of any one of the genes in Figure 16 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF6. Similarly, the invention provides a method of regulating the expression level of any one of the genes in Figure 17 in a pancreatic cell, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF6.

The invention additionally provides a method of regulating the expression level of any one of the genes in Figure 18 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF4alpha. Similarly, the invention provides a method of regulating the expression level of any one of the genes in Figure 19 in a pancreatic cell, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF4alpha.

Agents which modulate the transcriptional activity of HNF-4a, or any other HNF family member, may be identified by screening compounds for their ability to

increase the expression level, the DNA binding activity or the transcriptional promoting activity of HNF4a . One assay format which can be used employs two genetic constructs. One is typically a plasmid that continuously expresses the transcriptional regulator of interest when transfected into an appropriate cell line. CV-1 cells are most often used. The second is a plasmid which expresses a reporter, e.g., luciferase under control of the transcriptional regulator. For example, if a compound which acts as a ligand for HNF-4 is to be evaluated, one of the plasmids would be a construct that results in expression of the HNF-4 receptor in an appropriate cell line, e.g., the CV-1 cells. The second would possess a promoter linked to the luciferase gene in which an HNF-4 response element is inserted. If the compound to be tested is an agonist for the HNF-4 receptor, the ligand will complex with the receptor and the resulting complex binds the response element and initiates transcription of the luciferase gene. In time the cells are lysed and a substrate for luciferase added. The resulting chemiluminescence is measured photometrically. Dose response curves are obtained and can be compared to the activity of known ligands. Other reporters than luciferase can be used including CAT and other enzymes.

Viral constructs can be used to introduce the gene for the receptor and the reporter. An usual viral vector is an adenovirus. For further details concerning this preferred assay, see U.S. Pat. No. 4,981,784 issued Jan. 1, 1991 hereby incorporated by reference, and Evans et al., WO88/03168 published on 5 May 1988, also incorporated by reference.

HNF-4a antagonists can be identified using this same basic "agonist" assay. A fixed amount of an antagonist is added to the cells with varying amounts of test compound to generate a dose response curve. If the compound is an antagonist, expression of luciferase is suppressed.

Additional methods for the isolation of agonists and antagonist of HNF transcription factors are described in U.S. Patent Nos. 6,187,533 and 5,620,887. Additional U.S. patents describing methods to identify agents that modulate the activity of transcription factors include 5,804,374, and 5,298,429, and U.S. Patent Publication

Nos. 2004/0033942A1 2003/0077664, 2003/0215829 and 2003/0039980. Any of the methods described herein may be easily adapted to identify agonists or antagonists of any one of the HNF transcriptional factors. U.S. Patent No. 6,303,653 describes modulators of HNF-4 activity.

Agonists and antagonists of HNF4a can also be designed based on the known crystal structure of HNF4a complexed with an endogenous fatty acid ligand (Dhe-Paganon, J. Biol. Chem. 277(41), 37973-37976). U.S. Patent Publication No. 2002/0072587 describes methods of identifying agonists of an estrogen receptor, a nuclear receptor like the HNF proteins, based on its crystal structure. Such methods may easily be applied to HNF-1a, HNF-4a and HNF6 by one skilled in the art. Additional examples of rational drug design based on the structure of a protein may be found in U.S. Patent or Publication Nos. 6,236,946, 6,684,162, 2004/0014153, 2003/0124699 , 20030077628, 2002/0151028, 2002/0072587 and 2003/0211588.

6. Therapeutics

In one aspect, the invention provides methods of treating disease in a subject comprising the administration of a composition comprising a therapeutic agent. "Therapeutic agent" or "therapeutic" refers to an agent capable of having a desired biological effect on a host. Chemotherapeutic and genotoxic agents are examples of therapeutic agents that are generally known to be chemical in origin, as opposed to biological, or cause a therapeutic effect by a particular mechanism of action, respectively. Examples of therapeutic agents of biological origin include growth factors, hormones, and cytokines. A variety of therapeutic agents are known in the art and may be identified by their effects. Certain therapeutic agents are capable of regulating cell proliferation and differentiation. Examples include chemotherapeutic nucleotides, drugs, hormones, non-specific (non-antibody) proteins, oligonucleotides (e.g., antisense oligonucleotides that bind to a target nucleic acid sequence (e.g., mRNA sequence)), peptides, and peptidomimetics.

In one embodiment, the compositions are pharmaceutical compositions. Pharmaceutical compositions for use in accordance with the present invention may be

formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by, for example, by aerosol, intravenous, oral or topical route. The administration may comprise intralesional, intraperitoneal, subcutaneous, intramuscular or intravenous injection; infusion; liposome-mediated delivery; topical, intrathecal, gingival pocket, per rectum, intrabronchial, nasal, transmucosal, intestinal, oral, ocular or otic delivery.

An exemplary composition of the invention comprises an compound capable of modulating the expression or activity of a transcriptional regulator with a delivery system, such as a liposome system, and optionally including an acceptable excipient. In a preferred embodiment, the composition is formulated for injection.

Techniques and formulations generally may be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the compounds of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the compounds may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid

preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., ationd oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound. For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner. For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such

as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. in addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art. A wash solution can be used locally to treat an injury or inflammation to accelerate healing.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

For therapies involving the administration of nucleic acids, the oligomers of the invention can be formulated for a variety of modes of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, intranodal, and subcutaneous for injection, the oligomers of the invention can be formulated in liquid solutions, preferably in

physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics. For topical administration, oligomers may be formulated into ointments, salves, gels, or creams as generally known in the art.

Toxicity and therapeutic efficacy of the agents and compositions of the present invention can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic induces are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially

from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

In one embodiment of the methods described herein, the effective amount of the agent is between about 1mg and about 50mg per kg body weight of the subject. In one embodiment, the effective amount of the agent is between about 2mg and about 40mg per kg body weight of the subject. In one embodiment, the effective amount of the agent is between about 3mg and about 30mg per kg body weight of the subject. In one embodiment, the effective amount of the agent is between about 4mg and about 20mg per kg body weight of the subject. In one embodiment, the effective amount of the agent is between about 5mg and about 10mg per kg body weight of the subject.

In one embodiment of the methods described herein, the agent is administered at least once per day. In one embodiment, the agent is administered daily. In one embodiment, the agent is administered every other day. In one embodiment, the agent is administered every 6 to 8 days. In one embodiment, the agent is administered weekly.

As for the amount of the compound and/or agent for administration to the subject, one skilled in the art would know how to determine the appropriate amount. As used herein, a dose or amount would be one in sufficient quantities to either inhibit the disorder, treat the disorder, treat the subject or prevent the subject from becoming afflicted with the disorder. This amount may be considered an effective amount. A person of ordinary skill in the art can perform simple titration experiments to determine what amount is required to treat the subject. The dose of the composition of the invention will vary depending on the subject and upon the particular route of administration used. In one embodiment, the dosage can range from about 0.1 to about 100,000 ug/kg body weight of the subject. Based upon the composition, the dose can be

delivered continuously, such as by continuous pump, or at periodic intervals. For example, on one or more separate occasions. Desired time intervals of multiple doses of a particular composition can be determined without undue experimentation by one skilled in the art.

The effective amount may be based upon, among other things, the size of the compound, the biodegradability of the compound, the bioactivity of the compound and the bioavailability of the compound. If the compound does not degrade quickly, is bioavailable and highly active, a smaller amount will be required to be effective. The effective amount will be known to one of skill in the art; it will also be dependent upon the form of the compound, the size of the compound and the bioactivity of the compound. One of skill in the art could routinely perform empirical activity tests for a compound to determine the bioactivity in bioassays and thus determine the effective amount. In one embodiment of the above methods, the effective amount of the compound comprises from about 1.0 ng/kg to about 100 mg/kg body weight of the subject. In another embodiment of the above methods, the effective amount of the compound comprises from about 100 ng/kg to about 50 mg/kg body weight of the subject. In another embodiment of the above methods, the effective amount of the compound comprises from about 1 ug/kg to about 10 mg/kg body weight of the subject. In another embodiment of the above methods, the effective amount of the compound comprises from about 100 ug/kg to about 1 mg/kg body weight of the subject.

As for when the compound, compositions and/or agent is to be administered, one skilled in the art can determine when to administer such compound and/or agent. The administration may be constant for a certain period of time or periodic and at specific intervals. The compound may be delivered hourly, daily, weekly, monthly, yearly (e.g. in a time release form) or as a one time delivery. The delivery may be continuous delivery for a period of time, e.g. intravenous delivery. In one embodiment of the methods described herein, the agent is administered at least once per day. In one embodiment of the methods described herein, the agent is administered daily. In one embodiment of the methods described herein, the agent is administered every other day. In one embodiment of the methods described herein, the agent is administered every 6

to 8 days. In one embodiment of the methods described herein, the agent is administered weekly.

EXEMPLIFICATION

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention, as one skilled in the art would recognize from the teachings hereinabove and the following examples, that other DNA microarrays, transcriptional regulators, cell types, antibodies, ChIP conditions, or data analysis methods, all without limitation, can be employed, without departing from the scope of the invention as claimed.

The practice of the present invention will employ, where appropriate and unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, virology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are described in the literature. See, for example, Molecular Cloning: A Laboratory Manual, 3rd Ed., ed. by Sambrook and Russell (Cold Spring Harbor Laboratory Press: 2001); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Using Antibodies, Second Edition by Harlow and Lane, Cold Spring Harbor Press, New York, 1999; Current Protocols in Cell Biology, ed. by Bonifacino, Dasso, Lippincott-Schwartz, Harford, and Yamada, John Wiley and Sons, Inc., New York, 1999; and PCR Protocols, ed. by Bartlett et al., Humana Press, 2003.

Various publications, patents, and patent publications are cited throughout this application the contents of which are incorporated herein by reference in their entirety.

Experimental procedures

The following procedures were followed in performing the experiments below:

Genome-scale Location Analysis

The protocol described here was adapted from Ren 2001. Briefly, cells are fixed with1% final concentration formaldehyde for 10-20 minutes at room temperature, harvested and rinsed with 1x PBS. The resultant cell pellet is sonicated, and DNA fragments that are crosslinked to a protein of interest are enriched by immunoprecipitation with a factor specific antibody. After reversal of the crosslinking, the enriched DNA is amplified using ligation-mediated PCR (LM-PCR), and then fluorescently labeled using high concentration Klenow polymerase and a dNTP-fluorophore. A sample of DNA that has not been enriched by immunoprecipitation is subjected to LM-PCR and labeled with a different fluorophore. Both IP-enriched and unenriched pools of labeled DNA are hybridized to a single DNA microarray containing 13,000 human intergenic regions (see below for description of DNA microarray and binding site determination).For hepatocyte experiments, 2.5 x 107 hepatocytes were typically used per chromatin immunoprecipitation. These hepatocytes were isolated by standard liver perfusion techniques, immediately crosslinked with 1% formaldehyde solution, rinsed, and flash frozen. Islet preparations were treated with formaldehyde between 1 hour and 5 days after isolation from pancreata. A minimum of 30,000 viable islet equivalents (approximately 2x 10^7 beta cells) were fixed and handled as described above. Typical islet purity for three experiments described here was >70% islets with >80% viability. HNF4a, HNF6, and RNA polymerase II produced high quality results with as few as 30,000 islet equivalents. HNF1a ChIP required significantly more material, typically 80,000 islets, to produce results with somewhat lower enrichment ratios than the results obtained with hepatocytes.

Human 13K DNA Microarray

It would be ideal to have a DNA microarray that contains the entire human genome sequence, but technical limitations and cost led applicants to select the most relevant portion of the genome for inclusion in this microarray. Because a significant percentage of transcriptional binding sites in proximal promoters are within 1 kb of transcription start sites, applicants designed primers to amplify these genomic regions for printing onto a promoter array. Applicants selected 15000 cDNAs from the NCBI RefSeq database, and mapped them to NCBI Build 22 (April 2001) of the human

genome using BLAST. Where multiple splice variants had been described, applicants used the most upstream site, and verified the 5'-end by alignment with the Database of Transcriptional Start Sites (http://elmo.ims.utokyo.ac.jp/dbtss/). Sequences to be amplified were extracted from the genomic region–750 bp to +250 bp relative to this transcriptional start site. To control for nonspecific binding, 9 amplified regions derived from long Arabidopsis open reading frames were included on the array. As a further negative control and for use in data normalization, applicants chose 158 ORF regions within long exons of human genes for amplification. To prepare the DNA content of the arrays, the program Primer3 (http://wwwgenome.wi.mit.edu/genome_software/other/ primer3.html) was used to design primers using the sequences described above. PCRs were performed on these primer set using standard conditions, except for the presence of 1 M betaine in all PCR reactions. Betaine was empirically observed to increase the success rate of the amplification reactions.

Of the 13,000 PCR pairs, 70% gave a strong band of the appropriate size, as verified on 2% agarose gels. Applicants have noted, however, that PCR products undetectable by agarose EtBr gel analysis can give valid positive signals when concentrated and printed on the DNA arrays. PCR quality evaluations were performed on the BRIDNAsuite of programs from the Biotechnology Research Institute of the National Research Council of Canada (http://www.irb-bri.cnrc-nrc.gc.ca/).PCR products were recovered from the reaction mixture by ammonium acetate/isopropanol precipitation and resuspended into 3x SSC with 1.5 M betaine to minimize evaporation and improve spot quality. Applicants printed amplified products onto GAPS-coated glass slides (Corning) using a Cartesian PixSys 5500 arrayer. The quality of the arrays was determined on a batch-wise basis by hybridization with sequence neutral oligonucleotides covalently linked to Cy3 or Cy5, followed by calculation of usable percentage of spots, combined with direct visual inspection of the quality of the chip. The Hu13K array was remapped post-production using two independent methods. First, applicants performed electronic PCR on the primer sets against the August 2003 final release of the completed human genome. Second, applicants BLASTed the sequence used to extract primers for amplification against the August 2003 final release of the

human genome. The dataset downloadable from the supporting website reports the location of each arrayed promoter relative to the transcriptional start site.

Data Quality Control

1. ChIP-Hybridization Quality Control

The raw data generated from each array experiment was subjected to multiple levels of quality control. First, each scan was examined visually as it was being performed. Samples on microarrays with gross defects (e.g. scratches, smeared spots) were repeated whenever possible. Applicants also determined that no reliable signal was produced from control spots containing *Arabidopsis* DNA.

2. Binding Site Determination and Error Model

Scanned images were analyzed using GenePix (v3.1 or v4.0), to obtain background subtracted intensity values. Each spot is bound by both IP-enriched and unenriched DNA, which are labeled with different fluorophores. Consequently, each spot yields fluorescence intensity information in two channels, corresponding to immunoprecipitated DNA and genomic DNA. To account for background hybridization to slides, the median intensity of a set of control blank spots was subtracted for site-specific transcription factors (e.g. HNF1a), and the median intensity for a set of control ORF spots was subtracted for broadly acting DNA binding proteins (e.g. RNA Pol II, HNF4a). To correct for different amounts of genomic and immunoprecipitated DNA hybridized to the microarray, the median intensity value of the IP-enriched DNA channel was divided by the median of the genomic DNA channel, and this normalization factor was applied to each intensity in the genomic DNA channel. Next, applicants calculated the log of the ratio of intensity in the IP-enriched channel to intensity in the genomic DNA channel for each intergenic region across the entire set of hybridization experiments. Adjusted intensity values for the IP-enriched channel were calculated from these ratios. A whole-chip error model (Hughes 2000; Lee 2002) was then used to calculate confidence values for each spot on each microarray, and to combine data for the replicates of each experiment to obtain a final average ratio and confidence for each promoter region. Genes were included in the set of 'bound' genes if the binding P-value in the error model was < 0.001 or enrichment was at least 2-fold

in the immunoprecipitation.

Confirmation of Predicted Binding

The accuracy of genome-wide location data reported here has been assessed using several approaches.

1. Estimation of False Positive Rates Using Conventional ChIP Experiments

Conventional, independent ChIP experiments conducted in our laboratory at a gene specific level have confirmed over 100 binding interactions identified by location analysis data involving 6 different regulators (see http://web.wi.mit.edu/young/pancregulators). These results suggest that our empirical rate of false positives is at most 16%. This rate is somewhat higher than that found for a large scale survey of yeast transcription factors (Lee 2002), which probably reflects the greater complexity of the human genome. Figures 9 and 10 show typical verification ChIP experiments for HNF4a and HNF1a, respectively, in hepatocytes.

2. Comparison with Previous Literature

Applicants found no previous studies of the genomic targets of transcriptional regulators in primary human tissue. However, a large number of HNF1a and HNF4a targets have been identified in model organisms and human carcinoma (mostly hepatoma) cell lines; these targets are summarized in Figure 14. For example, genome-scale location analysis identified 30 of the 68 hepatocyte genes which were both previously suggested to be targets of HNF4a, and included on the 13K DNA array. Similarly, genome-scale location analysis identified 21 of the 81 hepatocyte genes which were both previously suggested to be targets of HNF4a, and included on the 13K DNA array. Discrepancies between the targets reported here and targets reported in the literature may result from a number of factors, which include, but are not limited to: (1) the limitations of using a 1 kb promoter fragment to probe the binding of a transcription factor, (2) the stringency of our threshold criteria, (3) the differences between the regulatory network in model organisms and/or cell lines, and the regulatory network in primary human tissue, (4) differences between indirect technologies in the literature (i.e. gel-shift and transient transfections) and genome-scale location analysis, (5) tissue isolation effects, among others. A more comprehensive discussion can be found at

http://web.wi.mit.edu/young/pancregulators

Regulatory Motifs Derived from Binding Data

In order to discover network motifs, two data matrices were created. The overall matrix D consists of binary entries Dij, where a 1 indicates binding of regulator j to intergenic region i, a 0 indicates no binding event. The regulator matrix R is a subset of D, containing only the rows corresponding to the intergenic region assigned to each regulator, in the same order as the columns of regulators. All analyses were performed in Matlab. The algorithms used to find each motif are described below. Autoregulatory motif: Find each non-zero entry on the diagonal of R. Feedforward loop: For each master regulator (column of R), find non-zero entries, which correspond to regulators bound. For each master regulator / secondary regulator pair, find all rows in D bound by both regulators. Multi-component loop: For each regulator (column of R), find the regulators to which it binds. For each of these, find the regulators it binds. If any of these are the original regulator, you have a multi-component loop of two. For all others, find regulators to which they bind. If any of these are the original, you have a multicomponent loop of three. Repeat to find larger loops. Single input module: Find the intergenic regions bound by only one regulator. That is, take the subset of rows of D such that the sum of each row is 1. Then for each regulator (column), find non-zero entries. Each set (greater than three intergenic regions) is a SIM. Multi-input module: Find the intergenic regions bound by more than one regulator. That is, take the subset of rows of D such that the sum of each row is greater than 1. Then, for each row, find any other row bound by the same regulators. The collection of rows bound by the same regulators correspond to a MIM. Once a row is assigned to a MIM, remove it from further analysis. Regulator chain: For each regulator (column of R), use a recursive algorithm to find chains of all lengths. That is, for each regulator whose promoter is bound by the regulator before it in the chain, find the regulator promoters to which it binds. Repeat until the chain ends. There are three possible ways to end a chain: a regulator that does not bind to the promoter of any other regulator, a regulator that binds to its own promoter, or one that binds to the promoter of another regulator earlier in the chain.

Example 1

The liver and pancreas have long been the subject of studies to understand how organs develop and are regulated at the transcriptional level (*8-12*). The transcriptional regulators HNF1α (a homeodomain protein), HNF4α (a nuclear receptor) and HNF6 (a member of the onecut family) operate cooperatively in a connected network in the liver, but less in known about the structure of this regulatory network in human pancreatic islets. All three transcriptional regulators are required for normal function of liver and pancreatic islets (*13-18*). Mutations in HNF1α and HNF4α are the causes of the type 3 and type 1 forms of maturity-onset diabetes of the young (MODY3 and MODY1), a genetic disorder of the insulin-secreting pancreatic beta cells characterized by onset of diabetes mellitus before 25 years of age and an autosomal dominant pattern of inheritance (19).

Applicants hypothesized that genome-scale analysis of the pancreatic islet genes whose expression is regulated by these transcription factors in normal beta cells could provide insights into the molecular basis of the abnormal beta cell function that characterizes MODY. Applicants have identified the genes occupied by the transcription factors HNF1α, HNF4α, and HNF6 in pancreatic islets. The genes transcribed in each tissue were identified by determining the genomic occupancy of RNA polymerase II. Applicants used this information to begin to map the transcriptional regulatory circuitry in these tissues.

Applicants first used genome-scale location analysis (*20*) to identify the promoters bound by HNF1α in human hepatocytes and pancreatic islets isolated from tissue donors (Fig 1A). For each tissue, HNF1α-DNA complexes were enriched by chromatin immunoprecipitation in three separate experiments. Applicants constructed a custom DNA microarray containing portions of promoter regions of 13,000 human genes (Hu13K array). Applicants targeted the region spanning 700 bp upstream and 200 bp downstream of transcription start sites for the genes whose start sites are best characterized based on National Center for Biotechnology Information annotation (*20*). Although many enhancers are present at more distant locations, most known

transcription factor binding site sequences occur within these start-site proximal regions of promoters.

The results of these genome location experiments revealed that HNF1α is bound to at least 222 target genes in hepatocytes, representing 1.6% of the genes on the Hu13K array (Figure 11) (*20*). This result was verified with independent, conventional chromatin immunoprecipitation experiments, which suggest that the frequency of false positives in genome-scale location data with gene-specific regulators is no more than 16% when our threshold criteria were used (*20*). The genes applicants found to be occupied by HNF1α in primary human hepatocytes encode products whose functions represent a significant cross-section of hepatocyte biochemistry. The results confirm that HNF1α contributes to the transcriptional regulation of many of the central rate-limiting steps in gluconeogenesis and associated pathways. HNF1α also binds to genes whose products are central to normal hepatic function, including carbohydrate synthesis and storage, lipid metabolism (synthesis of cholesterol and apolipoproteins), detoxification (synthesis of cytochrome P450s) and synthesis of serum proteins (albumin, complements and coagulation factors).

Applicants next identified HNF1α target genes in human pancreatic islets (Figure 11) (*20*). HNF1α occupied the promoter regions of 106 genes (0.8% of the Hu13K array promoters) in islets, 30% of which were also bound by HNF1α in hepatocytes (Figure 1B). In islets, fewer chaperones and enzymes are bound by HNF1α than in hepatocytes, and the receptors and signal transduction machinery regulated by HNF1α vary between the two tissues.

HNF1α has been previously implicated in the regulation of many genes in hepatocytes and islets (*13, 16, 20* [Figure 15]). The direct genome binding data reported here confirmed many, but not all, of these genes. The difference may be due, at least in part, to our stringent criteria for binding in the genome-scale data, which enhances our confidence in the direct target genes identified by location analysis, but likely underestimates the actual number of targets in vivo. Furthermore, although the

proximal promoter regions printed on the array contain a significant number of transcription factor binding sequences, many genes are also regulated by more distal promoter elements and enhancers that are not present on the Hu13K array.

Applicants also identified the promoters bound by HNF6 in human hepatocytes and pancreatic islets using genome-scale location analysis (Fig 1B; Figures 16 and 17) (*20*). HNF6 was bound to at least 222 genes in hepatocytes and 189 genes in pancreatic islets, representing 1.7% and 1.4% of the promoters on the array, respectively. Approximately half of the promoters occupied by HNF6 were common to the two tissues, and included a number of important cell cycle regulators such as CDK2 (*20*).

Genome-scale location analysis revealed surprising results for HNF4α in hepatocytes and pancreatic islets (Fig 1B). The number of genes enriched in HNF4α chromatin immunoprecipitations was much larger than observed with typical site-specific regulators. HNF4α was bound to approximately 12% of the genes represented on the Hu13K DNA microarray in hepatocytes and 11% in pancreatic islets. No other transcription factor applicants have profiled in human cells has been observed to bind more than 2.5% of the promoter regions represented on the 13K array.

Six independent lines of evidence indicate that the HNF4α results are not due to poor antibody specificity or errors in the microarray analysis, and support the view that HNF4α is associated with an unusually large number of promoters in hepatocytes and pancreatic islets (*20*). First, essentially identical results were obtained with two different antibodies that recognize different portions of HNF4α. Second, Western blots showed that the HNF4α antibodies are highly specific. Third, applicants verified binding at over 50 randomly selected targets of HNF4α in hepatocytes by conventional gene-specific chromatin immunoprecipitation. Fourth, when antibodies against HNF4α were used for ChIP in control experiments with Jurkat, U937, and BJT cells (which do not express HNF4α), no more than 17 promoters were identified in each cell line by our criteria, which is well within the noise inherent in this system. Fifth, when pre-immune antibodies from rabbit and goat (the two different anti-HNF4α antibodies came from rabbit and goat) were used in control experiments in hepatocytes, the

number of targets identified was within the noise. Finally, if the HNF4α results are correct, then applicants would expect that the set of promoters bound by HNF4α should be largely a subset of those bound by RNA polymerase II in each tissue; applicants found that this is the case (see below). Applicants conclude that HNF4α is a widely acting transcription factor in these tissues, consistent with the observation that it is an unusually abundant, constitutively active transcription factor (*11*).

Applicants next identified the genes represented on the Hu13K microarray that are actively transcribed in hepatocytes and pancreatic islets, so the fraction of actively transcribed genes that are bound by HNF4α could be determined (Fig 2C). It is difficult to determine accurately the transcriptome of these tissues by profiling transcript levels with DNA microarrays. Transcript profiling requires a reference RNA population against which a tissue RNA population can be compared, and there are limitations to generating appropriate reference RNA. To circumvent this limitation, applicants exploited the fact that RNA polymerase II occupies the set of protein-coding genes that are actively transcribed in eukaryotic cells. Location analysis with RNA polymerase II antibodies can identify these actively transcribed genes (*7, 21*). Applicants found that 23% of the genes on the Hu13K array (2984 genes) were bound by RNA polymerase II in hepatocytes, and 19% (2426 genes) were bound by RNA polymerase II in islets (*20*). The sets of genes occupied by RNA polymerase II in hepatocytes and islets overlapped substantially (81% overlap, relative to islets), consistent with the relatedness of the two tissues (*22*). As expected, the majority of genes occupied by HNF4α in hepatocytes and pancreatic islets (80% and 73%, respectively) were also occupied by RNA polymerase II. Remarkably, of the genes occupied by RNA polymerase II, 42% (1262/2984) were bound by HNF4α in hepatocytes and 43% (1047/2426) were bound by HNF4α in islets (Fig 1C). By comparison, only 6% and 2% of RNA polymerase II enriched promoters were also bound by HNF1α in hepatocytes and islets, respectively.

Previous studies indicate that HNF1α, HNF4α, and HNF6 are at the center of a network of transcription factors that cooperatively regulate numerous developmental and metabolic functions in hepatocytes and islets (*9, 13, 15, 17*). Our systematic

analysis of the direct in vivo targets of these factors significantly expands our understanding of the regulatory network in primary human tissues (Fig 2A). A comparison of the regulatory network in these two tissues reveals that HNF1α, HNF4α, and HNF6 occupy the promoters of genes encoding a large population of transcription factors and cofactors in the two tissues (*20*). The precise set of transcription factor genes occupied by HNF1α, HNF4α, and HNF6, and the extent to which they are co-occupied by the HNF regulators, differed substantially between these two tissues.

The transcription factor binding data was used to identify regulatory network motifs, simple units of transcriptional regulatory network architecture that suggest mechanistic models (Fig 2B) (*4, 23*). Our data confirm previous reports that HNF1α and HNF4α occupy one another's promoters in both hepatocytes and islets, forming a multi-component loop (*24-26*). Multicomponent loops provide the capacity for feedback control and produce bistable systems that can switch between two alternate states (*23*). It has been suggested that the multicomponent loop present between HNF1α and HNF4α is responsible for stabilization of the terminal phenotype in pancreatic beta cells (*26*). Applicants also found that HNF6 serves as a master regulator for feedforward motifs in hepatocytes and pancreatic islets involving over 80 genes in each tissue (Figures 20 and 22). For example, in hepatocytes, HNF6 binds the HNF4α7 promoter, and HNF6 and HNF4α together bind *PCK1*, which encodes phosphoenolpyruvate carboxykinase, an enzyme key to gluconeogenesis (Fig 2B). A feedforward loop can act as a switch designed to be sensitive to sustained, rather than transient, inputs (*23*). HNF1α, HNF4α and HNF6 were also found to form multi-input motifs by collectively binding to sets of genes in hepatocytes and islets. This regulatory motif suggests coordination of gene expression through multiple input signals. Applicants also found that HNF6, HNF4α, and HNF1α form a regulator chain motif with THRA (NR1D1); regulator chain motifs represent the simplest circuit logic for ordering transcriptional events in a temporal sequence (*4, 23*). Additional examples of these regulatory motifs can be found in Figures 20 and 23 (*20*). Figures 20-24, panels A and B, show transcriptional regulators occupied by HNF transcription factors and their regulatory loops. Figures 4-10 show additional controls and data generated by the experiments described herein.

Our results suggest that the nuclear hormone receptor HNF4α contributes to regulation of a large fraction of the liver and pancreatic islet transcriptomes by binding directly to almost half of the actively transcribed genes. This likely explains why HNF4α is crucial for development and proper function of these tissues (*12-15, 17, 18*). Perhaps most importantly, our results suggest a mechanistic explanation for the recent discovery that polymorphisms in the islet-specific P2 promoter for the splice variant HNF4α7 can greatly increase the risk of type II diabetes (*27-30*). Applicants found that multiple HNF factors bind directly to the P2 promoter in primary, healthy human islets. Alterations in the binding sites for these factors could cause misregulation of HNF4α expression and thus its downstream targets, leading to beta cell malfunction and diabetes.

References for Experimental Section:

1. Roeder, R.G. *Cold Spring Harb Symp Quant Biol* 63, 201 (1998).
2. T.I. Lee, R.A. Young. *Annu Rev Genet* 34, 77 (2000).
3. G. Orphanides, D. Reinberg. *Cell* 108, 439 (2002).
4. T.I. Lee, et al. *Science* 298, 799 (2002).
5. B. Ren, et al. *Genes Dev* 16, 245 (2002).
6. A.S. Weinmann, et al. *Genes Dev* 16, 235 (2002).
7. Z. Li, et al. *Proc Natl Acad Sci U S A* 100, 8164 (2003).
8. E. Lai, J.E. Darnell, Jr. *Trends Biochem Sci* 16, 427 (1991).
9. C.J. Kuo, et al. *Nature* 355, 457 (1992).
10. M. Pontoglio, et al. *Cell* 84, 575 (1996).
11. F.M. Sladek, Seidel, S. D. in. *Nuclear Receptors and Genetic Disease.* T. P. Burris, Ed. (Academic Press, New York, 2001).
12. F. Parviz, et al. *Nat Genet* 34, 292 (2003).
13. R.H. Costa, et al, Hepatology 38, 1331 (2003).
14. D.Q. Shih, et al. *Diabetes* 50, 2472 (2001).
15. D.Q. Shih, M. Stoffel. *Proc Natl Acad Sci U S A* 98, 14189 (2001).
16. K.S. Zaret. *Nat Rev Genet* 3, 499 (2002).
17. P. Jacquemin et al. *Dev Biol* 258, 105 (2003).

18. Fajans, S. S., et al. *N Engl J Med* 345, 971 (2001).

19. See supporting data on Science Online, and additional information is available at the authors' website: http://web.wi.mit.edu/young/pancregulators

20. H.H. Ng, F. Robert, R.A. Young, K. Struhl. *Genes Dev* 16, 806 (2002).

21. R. Bort, K. Zaret. *Nat Genet* 32, 85 (2002).

22. R. Milo, et al. *Science* 298, 824 (2002).

23. S.F. Boj ae tl.*Proc Natl Acad Sci U S A* 98, 14481 (2001).

24. H. Thomas, et al. *Hum Mol Genet* 10, 2089 (2001).

25. J. Ferrer. *Diabetes* 51, 2355 (2002).

26. I. Barroso et al. *PLoS Biology* 1, 41 (2003).

27. Q. Zhu et al. *Diabetologia* 46, 567 (2003).

28. L. Love-Gregory et al. *Diabetes* 54 (2004) in press.

29. K. Silander et al. *Diabetes* 54 (2004) in press.

Claims:

1. A method of determining which genes from a subset of genes are regulated by a transcriptional regulator expressed in a cell, the method comprising

(a) selectively isolating chromatin from the cell to generate isolated chromatin;

(b) selectively isolating chromatin fragments from the isolated chromatin to generate bound chromatin fragments, wherein the bound chromatin fragments are bound by the transcriptional regulator;

(c) amplifying both the bound chromatin fragments to generate amplified chromatin fragments and the isolated chromatin to generate amplified control chromatin;

(d) hybridizing the amplified control chromatin and the amplified chromatin fragments to a DNA microarray, wherein the DNA microarray comprises

(1) at least 10,000 experimental spots, each experimental spot comprising an experimental DNA, each experimental DNA comprising a promoter region from a gene in the subset; and

(2) at least 100 control spots, each control spot comprising a control DNA, each control DNA comprising a non-promoter region; and

(e) determining and comparing a hybridization signal at each of the spots on the microarray between those generated by

(1) the amplified control chromatin; and

(2) the amplified chromatin fragments;

wherein a gene in the subset is said to be regulated by the transcriptional regulator in the cell if a spot comprising a promoter region of said gene displays a higher level of hybridization by the amplified chromatin fragments than by the amplified control chromatin.

2. The method of claim 1, wherein the level of hybridization of the amplified chromatin fragments to each experimental spot is normalized by the level of hybridization of the amplified chromatin fragments to the control spots.

3. The method of claim 1, wherein the level of hybridization of the amplified chromatin fragments to each experimental spot is normalized by subtracting the mean level of hybridization of the amplified chromatin fragments to the control spots.

4. The method of claim 1, wherein the higher level of hybridization comprises at least a two-fold higher level of hybridization.

5. The method of claim 1, wherein the transcriptional regulator is native to the cell.

6. The method of claim 1, wherein the transcriptional regulator is not a recombinant transcriptional regulator.

7. The method of claim 1, wherein the cell is a primary cell.

8. The method of claim 7, wherein the cell is a human cell.

9. The method of claim 8, wherein the cell is a transplant-grade human cell.

10. The method of claim 1, wherein step (b) comprises immunoprecipitation of the transcriptional regulator.

11. The method of claim 1, wherein step (c) comprises ligation-mediated polymerase chain reaction (LM-PCR).

12. The method of claim 1, wherein the promoter region of the gene comprises from at least 700bp upstream to at least 200 bp downstream of the transcriptional start site of the gene.

13. The method of claim 1, wherein the promoter region comprises at least 30, 40, 50, or 60 or nucleotides in length.

14. The method of claim 1, wherein the promoter region of the gene comprises a sequence of at least 30 nucleotides whose sequence is identical to a region stretching from 3 kb upstream to 1 kb downstream of the transcriptional start site of said gene.

15. The method of claim 1, wherein the non-promoter region comprises an open reading frame.

16. The method of claim 1, wherein the transcriptional regulator is a basal transcription factor.

17. The method of claim 16, wherein the transcriptional regulator is an RNA polymerase II or a TATA-binding protein.

18. A method of identifying a transcriptional regulatory network in a cell, the method comprising determining if a transcriptional regulator regulates additional transcriptional regulators in the cell using the method of claim 1, wherein a transcriptional regulatory network is identified if at least one additional transcriptional regulator is determined to be regulated by the transcriptional regulator.

19. The method of claim 18, wherein the experimental DNA comprises promoter regions from the additional transcriptional regulators.

20. A method of identifying a transcriptional regulatory network in a cell, the method comprising determining if a transcriptional regulator regulates

(i) its own promoter; or

(ii) a promoter from a plurality of transcriptional regulators,

using the method of claim 1, wherein the experimental DNA comprises

(a) a promoter from the transcriptional regulator; and

(b) promoters from the plurality of transcriptional regulators;

wherein a transcriptional regulatory network is identified if the transcriptional regulator regulates itself or if it regulates at least one of the plurality of transcriptional regulators.

21. A method of identifying transcriptional regulatory networks in a cell, the method comprising

(a) determining, by repeating the method of claim 1 for each of a plurality of transcriptional regulators, the genes in a subset which are regulated by each of the plurality of transcriptional regulators, wherein the experimental DNA comprises promoter regions for each of the plurality of transcriptional regulators;

(b) determining if any one of the plurality of transcriptional regulators are regulated by at least one of the plurality of transcriptional regulators;

wherein a transcriptional regulatory network is identified if any one of the plurality of transcriptional regulators is regulated by at least one of the plurality of transcriptional regulators.

22. The method of claim 21, further comprising determining if a gene is regulated by more than one of the plurality of transcriptional regulators.

23. A DNA microarray for determining promoter occupancy in a human cell, the microarray comprising

(1) at least 10,000 experimental spots, each experimental spot comprising an experimental DNA, each experimental DNA comprising a promoter region from a human gene in the subset; and

(2) at least 100 control spots, each control spot comprising a control DNA, each control DNA comprising a non-promoter region;

wherein at least 75% of the promoter regions comprise from at least 700bp upstream to at least 200 bp downstream of the transcriptional start site.

24. A method of estimating if a transcriptional regulator is a global transcriptional regulator, the method comprising

(a) selectively isolating chromatin from a tissue;

(b) identifying promoter regions from the chromatin which are bound by a candidate global transcriptional regulator;

(c) identifying promoter regions from the chromatin which are bound by a member of the basal transcriptional machinery; and

(d) comparing the promoter regions identified in steps (b) and (c) to determine the ratio between (i) the number of promoter regions bound by both the candidate global transcriptional regulator and the member of the basal transcriptional machinery; and (ii) the number of promoter regions bound by the member of the basal transcriptional machinery

wherein a transcriptional regulator is a global transcriptional regulator when the ratio is greater than 0.2.

25. The method of claim 24, wherein steps (b) and (c) are performed using a DNA microarray.

26. The method of claim 25, wherein the DNA microarray comprises

(i) at least 10,000 experimental spots, each experimental spot comprising an experimental DNA, each experimental DNA comprising a promoter region from a human gene in the subset; and

(ii) at least 100 control spots, each control spot comprising a control DNA, each control DNA comprising a non-promoter region;

27. The method of claim 24, wherein the member of the basal transcriptional machinery is an RNA polymerase II or a TATA-binding protein.

28. The method of claim 24, wherein the tissue is transplant-grade tissue.

29. The method of claim 24, wherein the tissue is freshly-isolated human tissue.

30. The method of claim 29, wherein the tissue is from a subject afflicted with a disorder.

31. The method of claim 30 wherein the disorder is a hyperplastic condition.

32. A method of identifying at least one target gene for the development of a therapeutic to treat or prevent a disorder in a subject, wherein at least one form of the disorder is caused by an altered activity in a transcriptional regulator or in a suspected transcriptional regulator, the method comprising

(a) identifying the genes regulated by the transcriptional regulator in a cell;

(b) determining if the transcriptional regulator is a broad-acting transcriptional regulator or a narrow-acting transcriptional regulator, wherein if the transcriptional regulator is a broad acting transcriptional regulator then the transcriptional regulator is a target gene for the development of a therapeutic, and wherein if the transcriptional regulator is a narrow acting transcriptional regulator then

(i) determining if at least one gene regulated by the transcriptional regulator is likely causative in the disorder, wherein a gene that is likely causative in the disorder is a target gene for the development of a therapeutic; and

(ii) reiterating steps (a) and (b) for at least one gene that is regulated by the transcriptional regulator in the cell and that either

(1) encodes a transcriptional regulator or

(2) is suspected to encode a transcriptional regulator,

with the modification that the transcriptional regulator of steps (a) and (b) is said gene,

thereby identifying at least one target gene for the development of a therapeutic to treat or prevent a disorder in the subject.

33. The method of claim 32, wherein identifying the genes regulated by the

transcriptional regulator in a cell comprises chromosome-wide location analysis.

34. The method of claim 32, wherein identifying the genes regulated by the transcriptional regulator in the cell comprises using the method of claim 1.

35. The method of claim 32, wherein the transcriptional regulator is a master regulatory gene.

36. The method of claim 35, wherein the master regulatory gene is SOX1-18, OCT6, PAX3, Myocardin, GATA1-6, TCF1/HNF1A, HNF4A, HNF6, NGN3, C/EBP, FOXA1-3, IPF1, GATA, HNF3, NKX2.1, CDX, FTF/NR5A2, C/EBPbeta, SCL1, SKIN1, or a member of the neurogenin, LK, LMO, SOX, OCT, PAX, GATA or MyoD family of transcription factors.

37. The method of claim 32, wherein the transcriptional regulator is PAX3, EGR-1, EGR-2, OCT6, a SOX family member, a GATA family member, a PAX family member, an OCT family member, RFX5, WHN, GATA1, VDR, CRX, CBP, MeCP2, AML1, p53, PLZF, PML, Rb, WT1, NR3C2, GCCR, PPARgamma, SIM1, HNF1alpha, HNF1beta, HNF4alpha, PDX1, MAFA, FOXA2, or NEUROD1.

38. The method of claim 32, wherein the cell is derived from a tissue whose function is impaired in the disorder.

39. The method of the claim 32, wherein the broad acting gene regulates at least about 2.5% of the genes in the cell, and wherein the narrow acting gene regulates less than about 2.5% of the genes in the cell.

40. The method of claim 32, wherein the gene is suspected to encode a transcriptional regulator if it shares at least 30% amino acid sequence identity with the DNA binding domain of a transcriptional regulator.

41. The method of claim 32, wherein the transcriptional regulator in the cell is a mutant transcriptional regulator.

42. The method of claim 32, wherein the transcriptional regulator in the cell has altered activity.

43. The method of claim 32, wherein the gene regulated by the transcriptional regulator is likely causative of the disorder when a mutation in the gene results in at least one phenotype or symptom associated with the disorder.

44. The method of claim 32, wherein the gene regulated by the transcriptional regulator is likely causative of the disorder when the gene encodes an enzyme or signaling molecule which functions in a pathway that is impaired in the disorder.

45. The method of claim 32, wherein the altered activity in the transcriptional regulator comprises at least one of the following:

(a) an alteration in the binding affinity of the transcriptional regulator to DNA;

(b) an alteration in the ability of the transcriptional regulator to bind to RNA polymerase, to an RNA polymerase holoenzyme, or to a second transcriptional regulator;

(c) an alteration in the binding affinity of the transcriptional regulator to a ligand;

(d) an alteration in expression level or expression pattern of the transcriptional regulator; or

(e) an alteration in an ability of the transcriptional regulator to form homomultimers or heteromultimers.

46. The method of claim 32, wherein the disorder is characterized by impaired function of at least one of the following: brain, spinal cord, heart, arteries,

esophagus, stomach, small intestine, large intestine, liver, pancreas, lungs, kidney, urinary tract, ovaries, breasts, uterus, testis, penis, colon, prostate, bone, muscle, cartilage, thyroid gland, adrenal gland, pituitary, bone marrow, blood, thymus, spleen, lymph nodes, skin, eye, ear, nose, teeth or tongue.

47. The method of claim 32, wherein the therapeutic comprises a small molecule drug, an antisense reagent, an antibody, a peptide, a ligand, a fatty acid, a hormone or a metabolite.

48. The method of claim 32, wherein the subject is a mammal.

49. The method of claim 48, wherein the mammal is a human.

50. The method of claim 32, wherein the transcriptional regulator is a transcriptional activator or a transcriptional repressor.

51. The method of claim 32, wherein the transcriptional regulator is native to the cell.

52. The method of claim 32, wherein the transcriptional regulator is from a species different from that of the cell.

53. The method of claim 52, wherein the transcriptional regulator is a viral transcriptional regulator.

54. A method of treating or preventing type II diabetes in a subject, comprising administering to the subject a therapeutically effective amount of an agent that increases the global transcriptional activity of HNF4alpha.

55. A method of treating or preventing a disorder associated with low transcriptional activity of HNF4alpha in a subject, comprising administering to the subject a therapeutically effective amount of an agent that increases the

global transcriptional activity of HNF4alpha.

56. A method of treating or preventing a disorder associated with high transcriptional activity of HNF4alpha in a subject, comprising administering to the subject a therapeutically effective amount of an agent that decreases the global transcriptional activity of HNF4alpha.

57. A method of increasing the global transcriptional activity in a liver or a pancreatic cell comprising contacting the cell with an agent which increases the global transcriptional activity of HNF4alpha.

58. A method of decreasing the global transcriptional activity in a liver or a pancreatic cell comprising contacting the cell with an agent which decreases the global transcriptional activity of HNF4alpha.

59. A method of regulating the expression level of any one of the genes in Figure 13 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF1alpha.

60. A method of regulating the expression level of any one of the genes in Figure 14 in a pancreatic cell, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF1alpha.

61. A method of regulating the expression level of any one of the genes in Figure 16 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF6.

62. A method of regulating the expression level of any one of the genes in Figure 17 in a pancreatic cell, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF6.

63. A method of regulating the expression level of any one of the genes in Figure

18 in a hepatocyte, the method comprising contacting the cell with an agent which regulates the transcriptional activity of HNF4alpha.

64. A method of regulating the expression level of any one of the genes in Figure 19 in a pancreatic cell, the method comprising contacting the cell with an agent which regulated the transcriptional activity of HNF4alpha.

65. A method of identifying transcriptionally active genes that are regulated by a transcriptional regulator in a cell, the method comprising

(a) selectively isolating chromatin from a tissue;

(b) identifying promoter regions from the chromatin that are bound by the transcriptional regulator;

(c) identifying promoter regions from the chromatin that are bound by a member of the basal transcriptional machinery; and

(d) comparing the promoter regions identified in steps (b) and (c) to determine overlapping genes,

wherein the overlapping genes are transcriptionally active genes regulated by the transcriptional regulator.

Fig. 1A

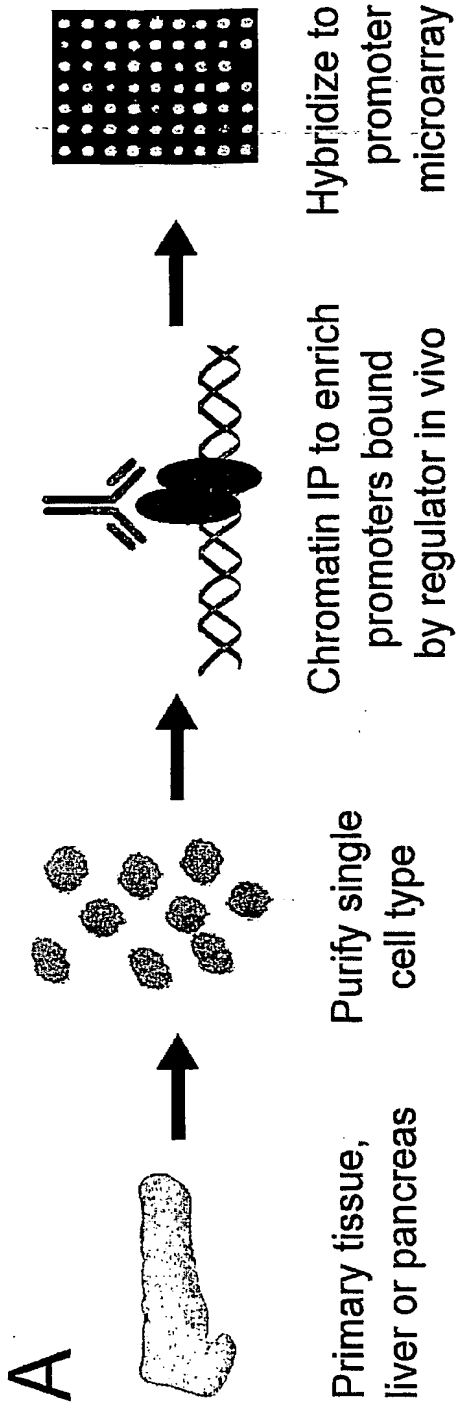

A
Primary tissue, liver or pancreas
Purify single cell type
Chromatin IP to enrich promoters bound by regulator in vivo
Hybridize to promoter microarray

Fig. 1B

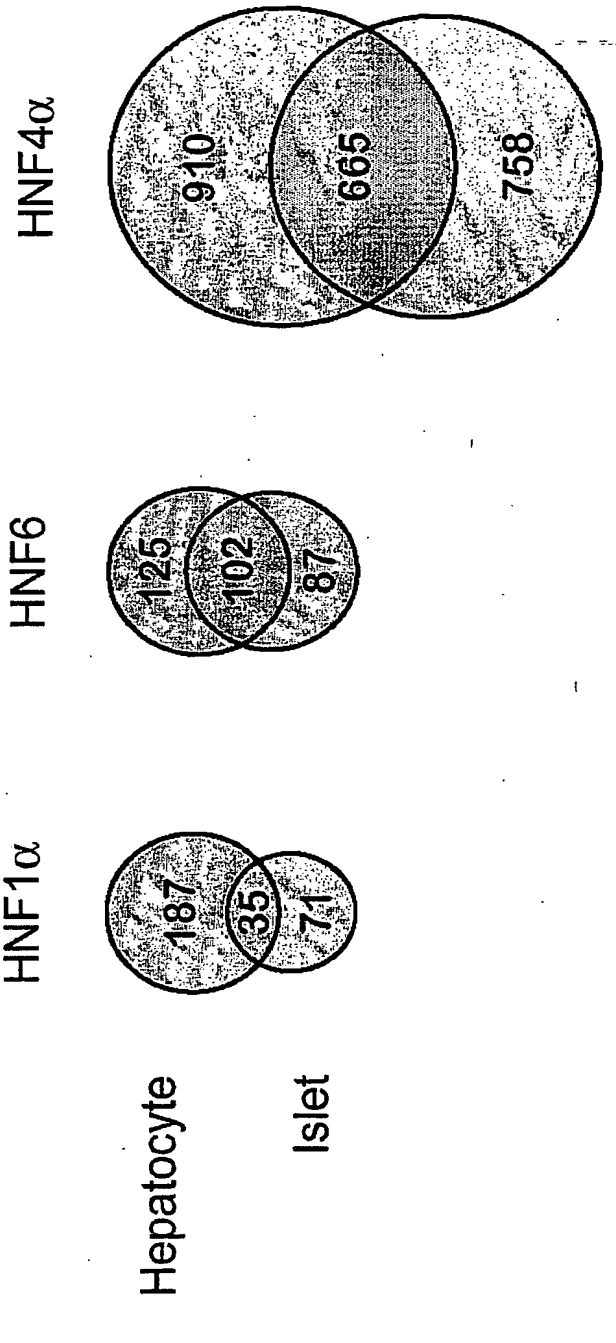

HNF4α
910
665
758
HNF6
125
102
87
HNF1α
187
35
71
Hepatocyte
Islet

Fig. 1C

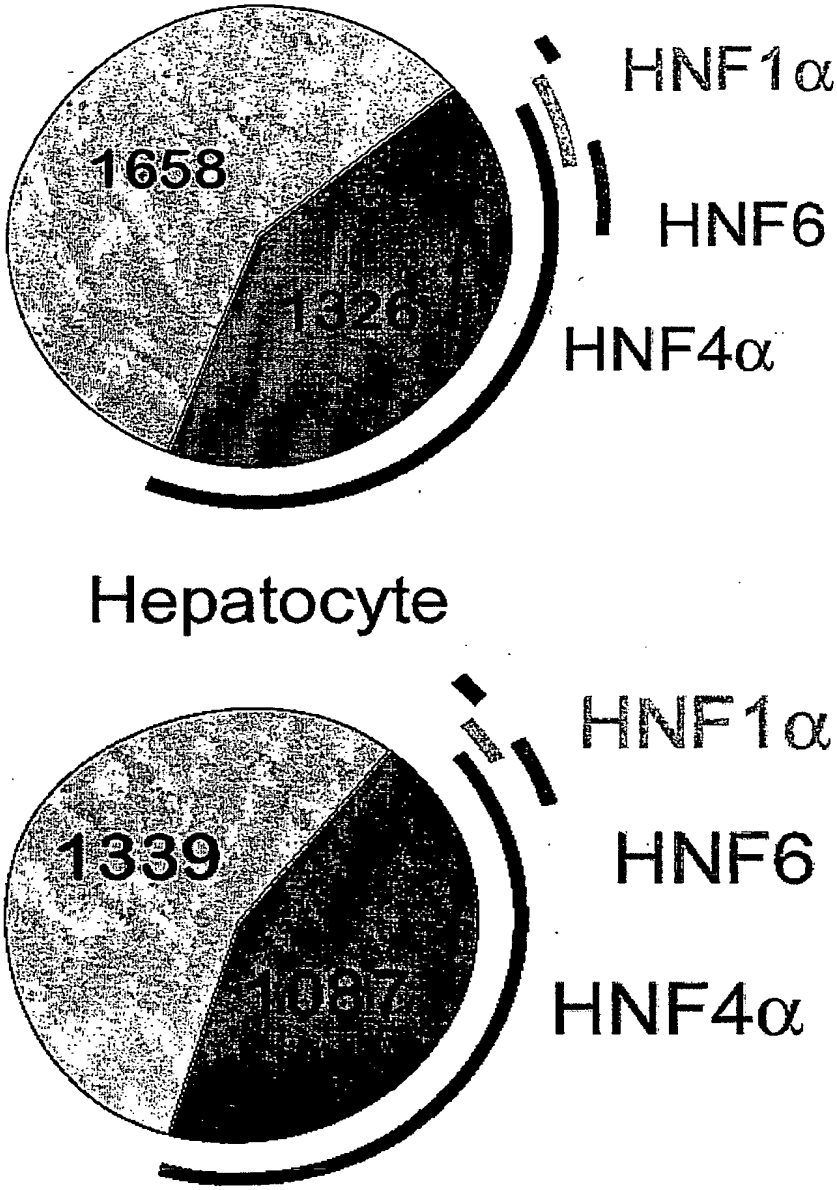

HNF1α
HNF6
HNF4α
1658
1326
Hepatocyte
HNF1α
HNF6
HNF4α
1339
1087
Pancreatic Islet

Fig. 2A

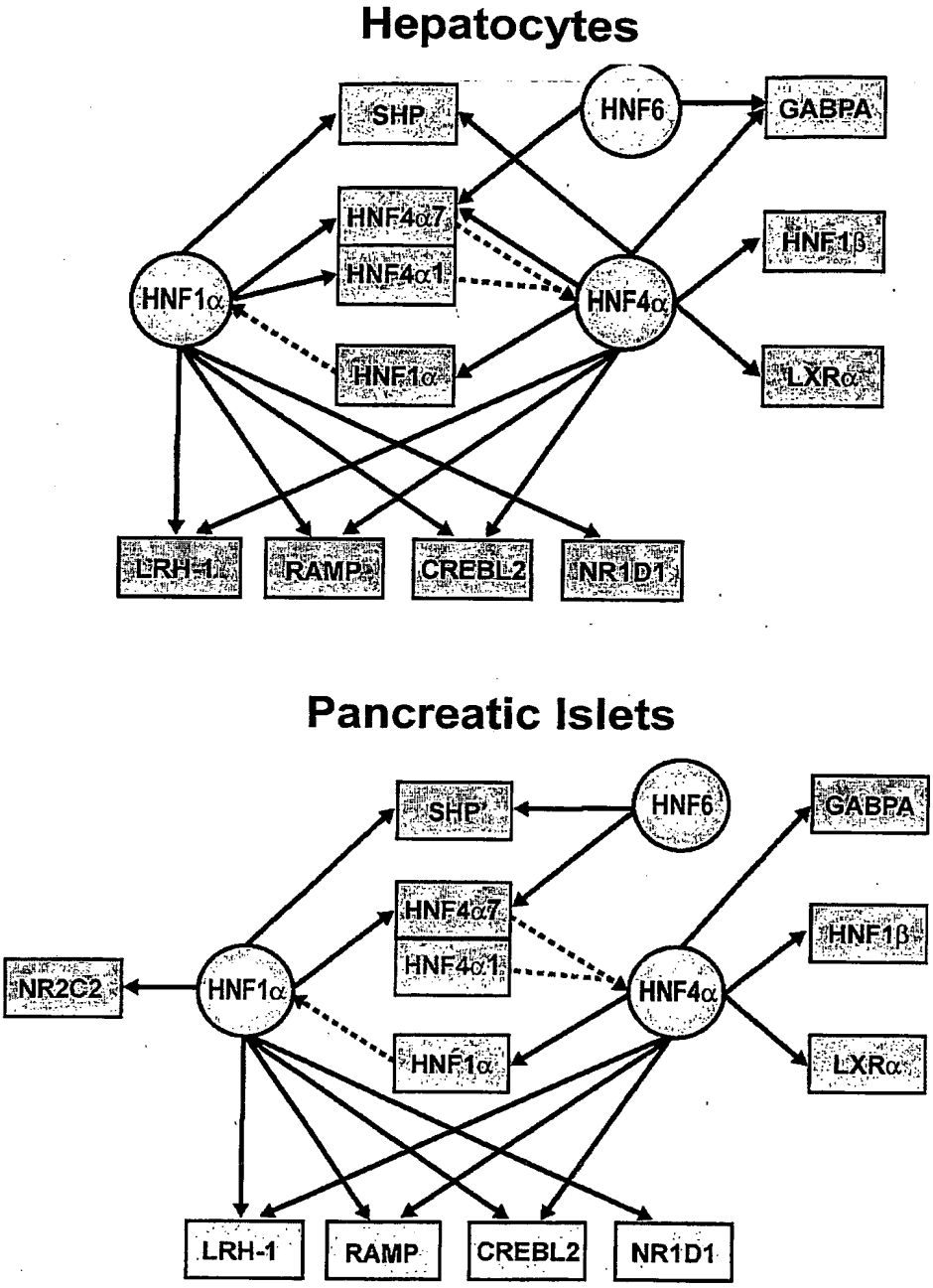

Hepatocytes
SHP
HNF6
GABPA
HNF4α7
HNF1β
HNF4α1
HNF1α
HNF4α
HNF1α
LXRα
LRH-1
RAMP
CREBL2
NR1D1
Pancreatic Islets
SHP
HNF6
GABPA
HNF4α7
HNF1β
NR2C2
HNF1α
HNF4α1
HNF4α
HNF1α
LXRα
LRH-1
RAMP
CREBL2
NR1D1

Fig. 2B

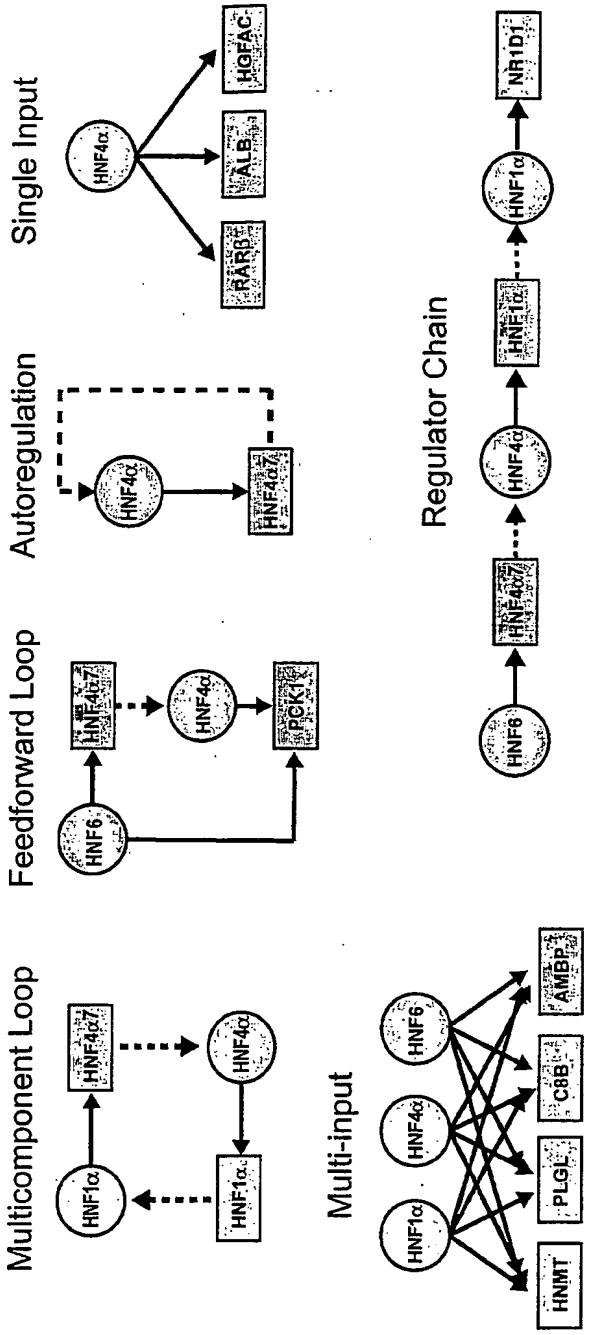

Single Input
HNF4α
HGFAC
ALB
RARβ
Autoregulation
HNF4α
HNF4α7
Feedforward Loop
HNF6
HNF4α7
HNF4α
PCK1
Regulator Chain
HNF6
HNF4α7
HNF4α
HNF1α
NR1D1
Multicomponent Loop
HNF1α
HNF4α7
HNF4α
HNF1α
Multi-input
HNF6
HNF4α
HNF1α
AMBP
C8B
PLGL
HNMT

Fig. 3

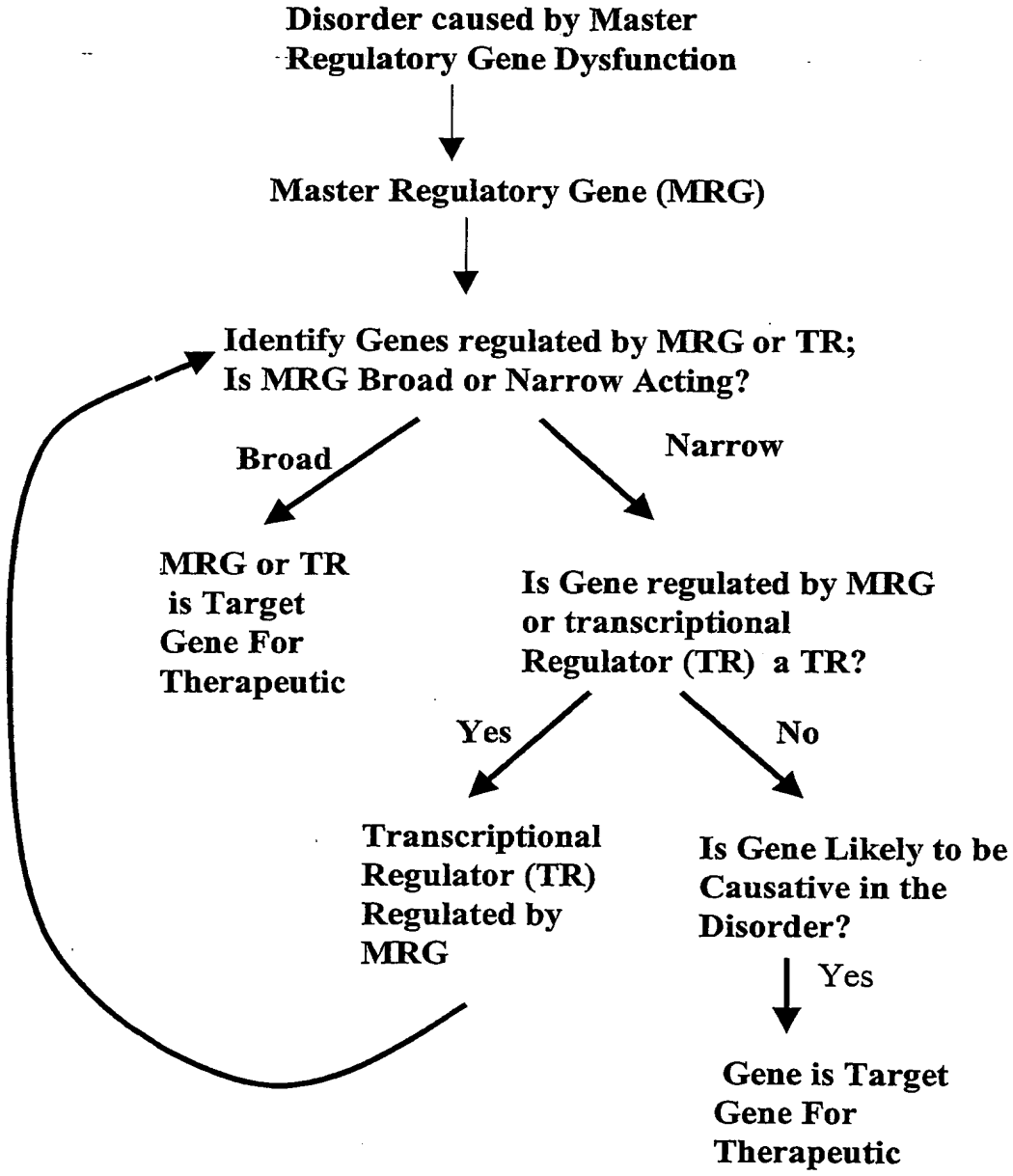

Disorder caused by Master Regulatory Gene Dysfunction
Master Regulatory Gene (MRG)
Identify Genes regulated by MRG or TR; Is MRG Broad or Narrow Acting?
Broad
Narrow
MRG or TR is Target Gene For Therapeutic
Is Gene regulated by MRG or transcriptional Regulator (TR) a TR?
Yes
No
Transcriptional Regulator (TR) Regulated by MRG
Is Gene Likely to be Causative in the Disorder?
Yes
Gene is Target Gene For Therapeutic

Fig. 4

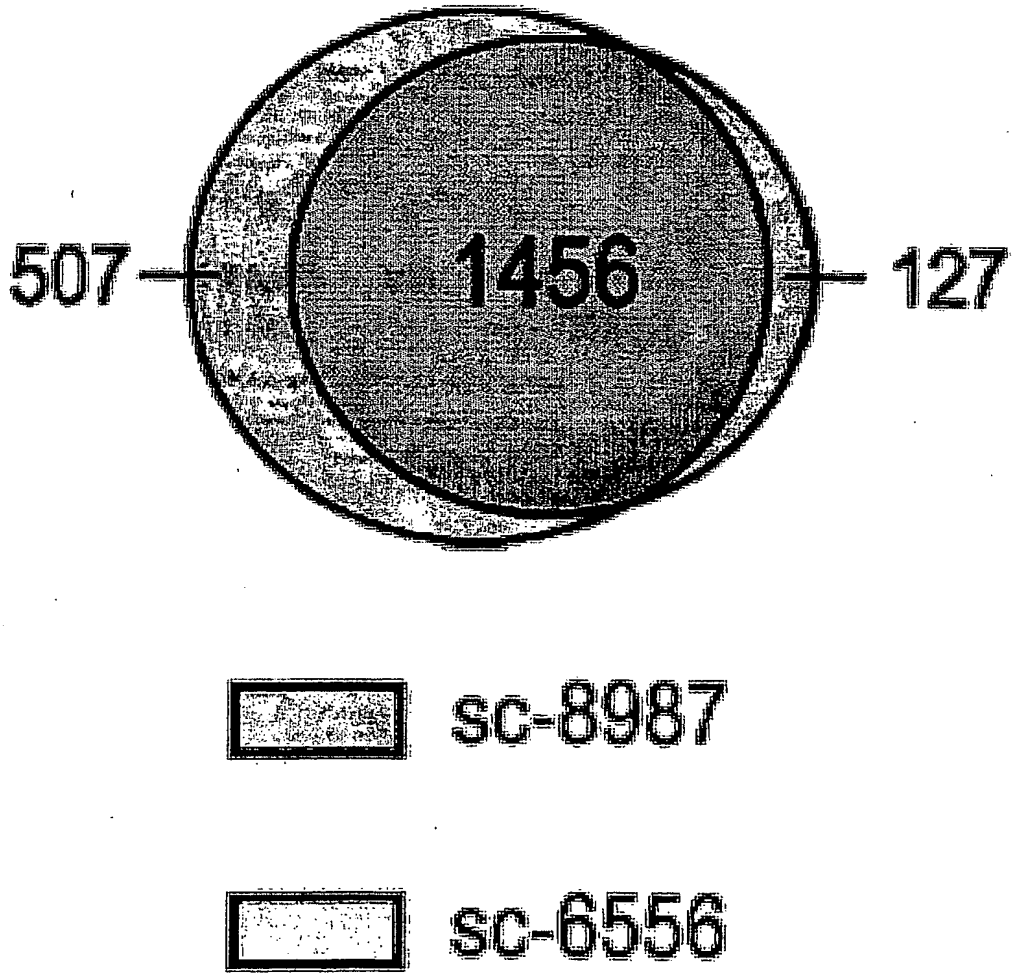


Fig. 5

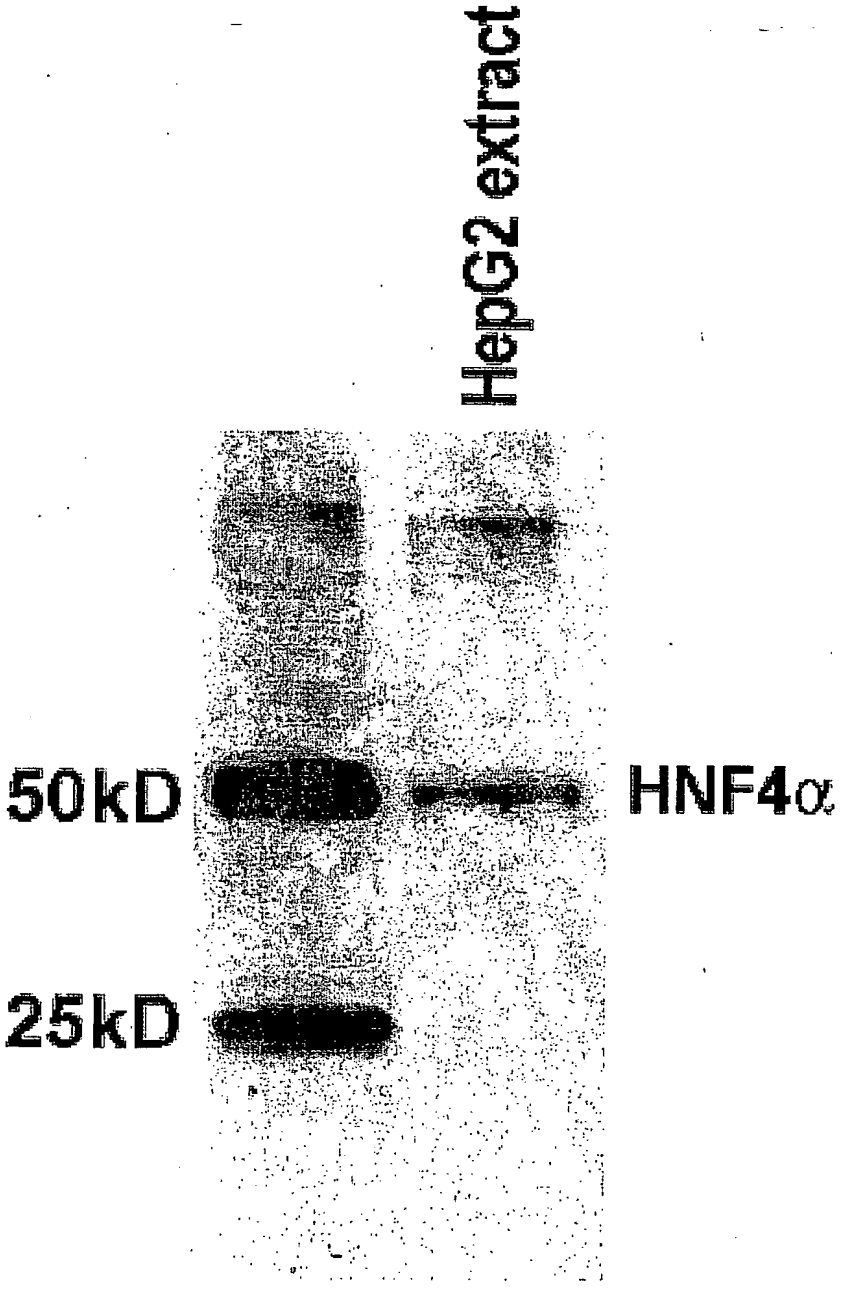


Fig. 6A

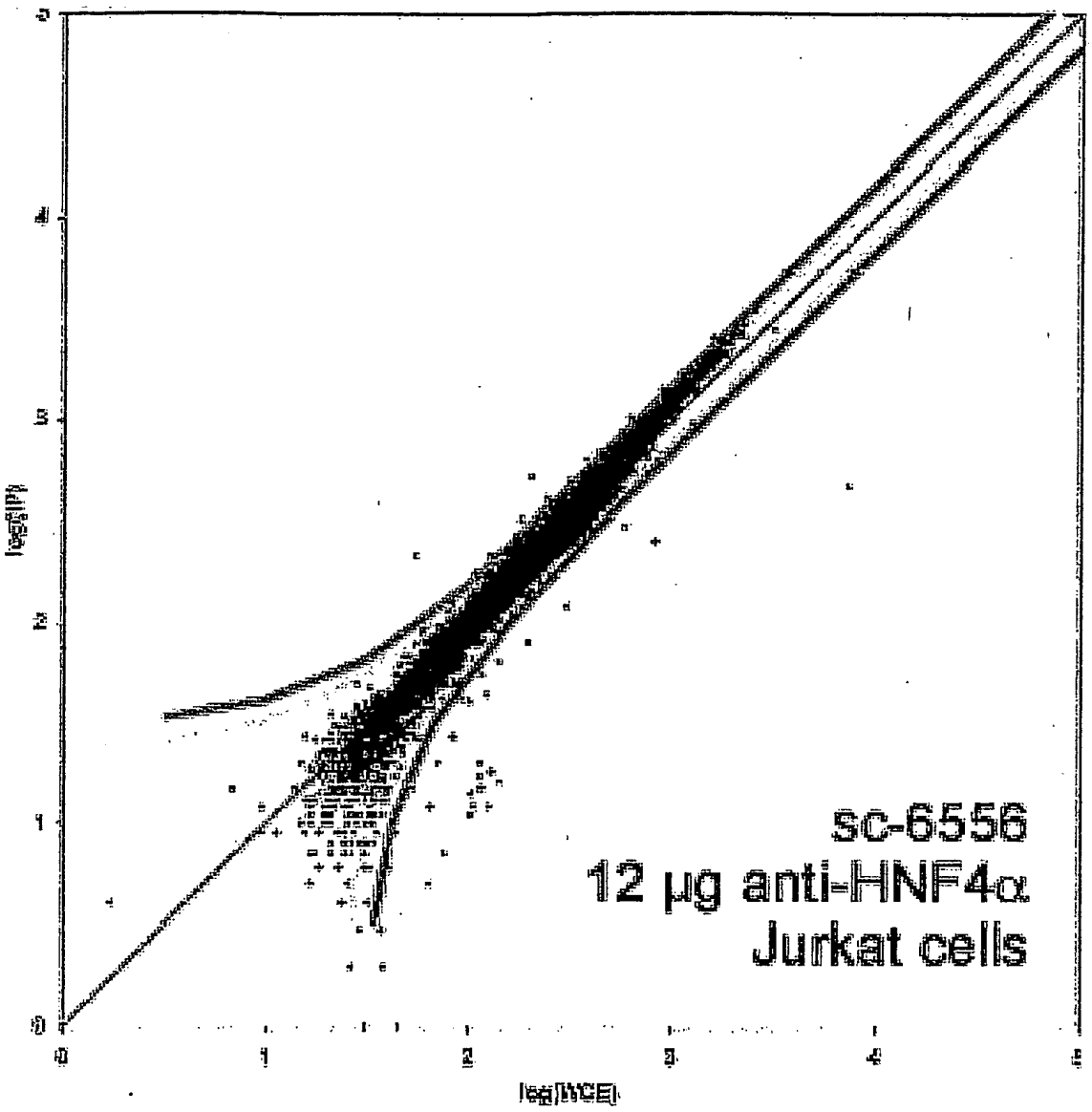


Fig. 6B

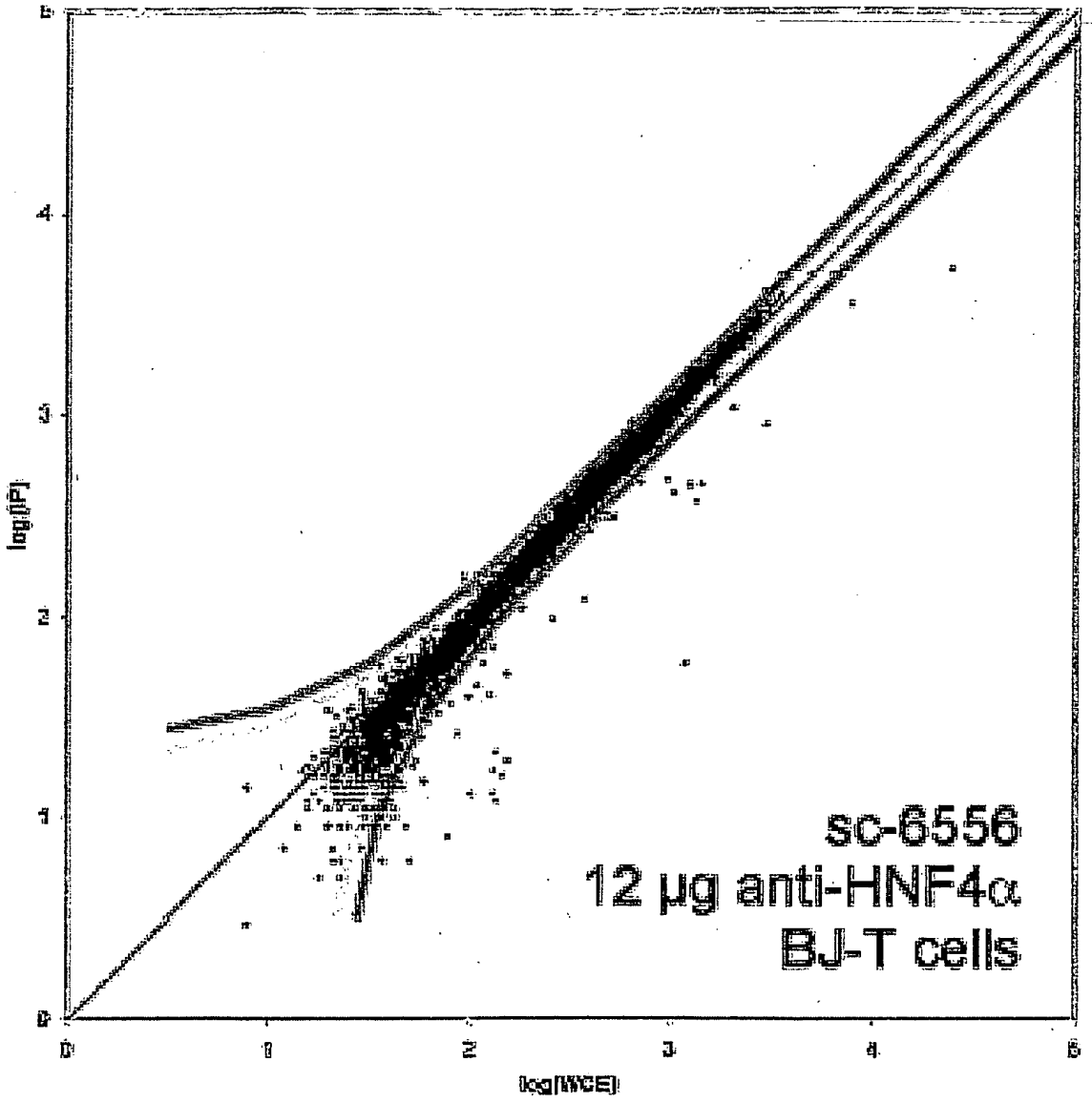


Fig. 6C

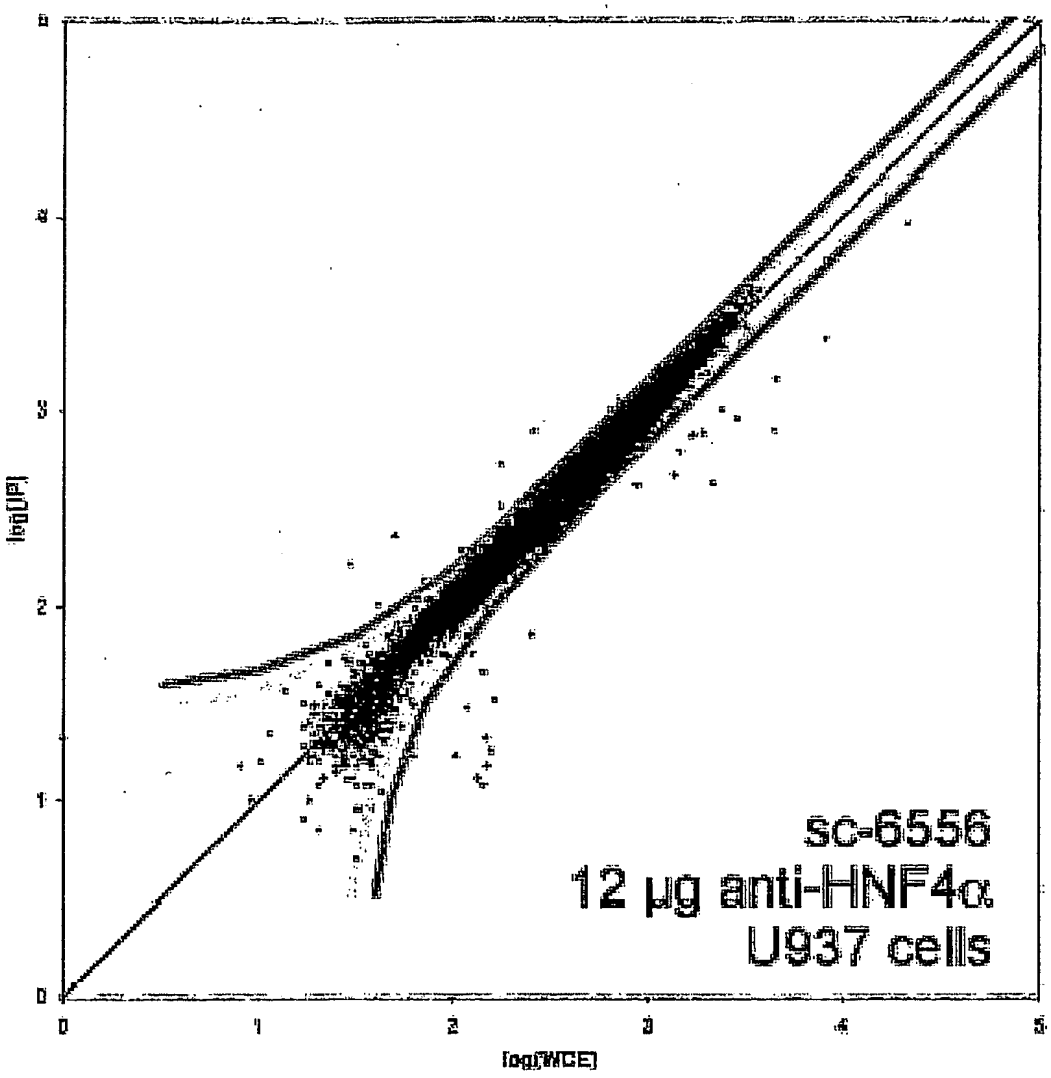


Fig. 6D

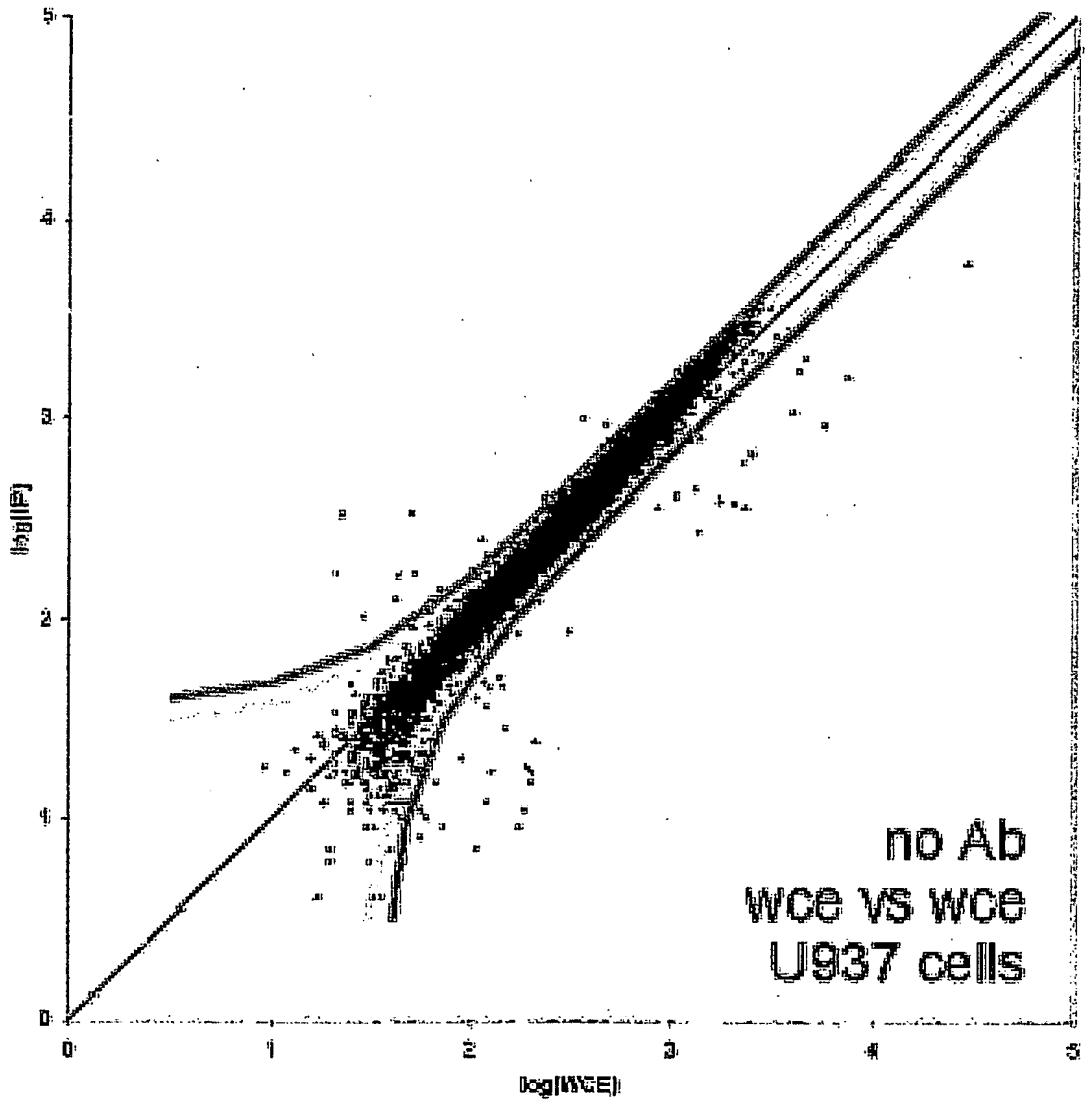


Fig. 7

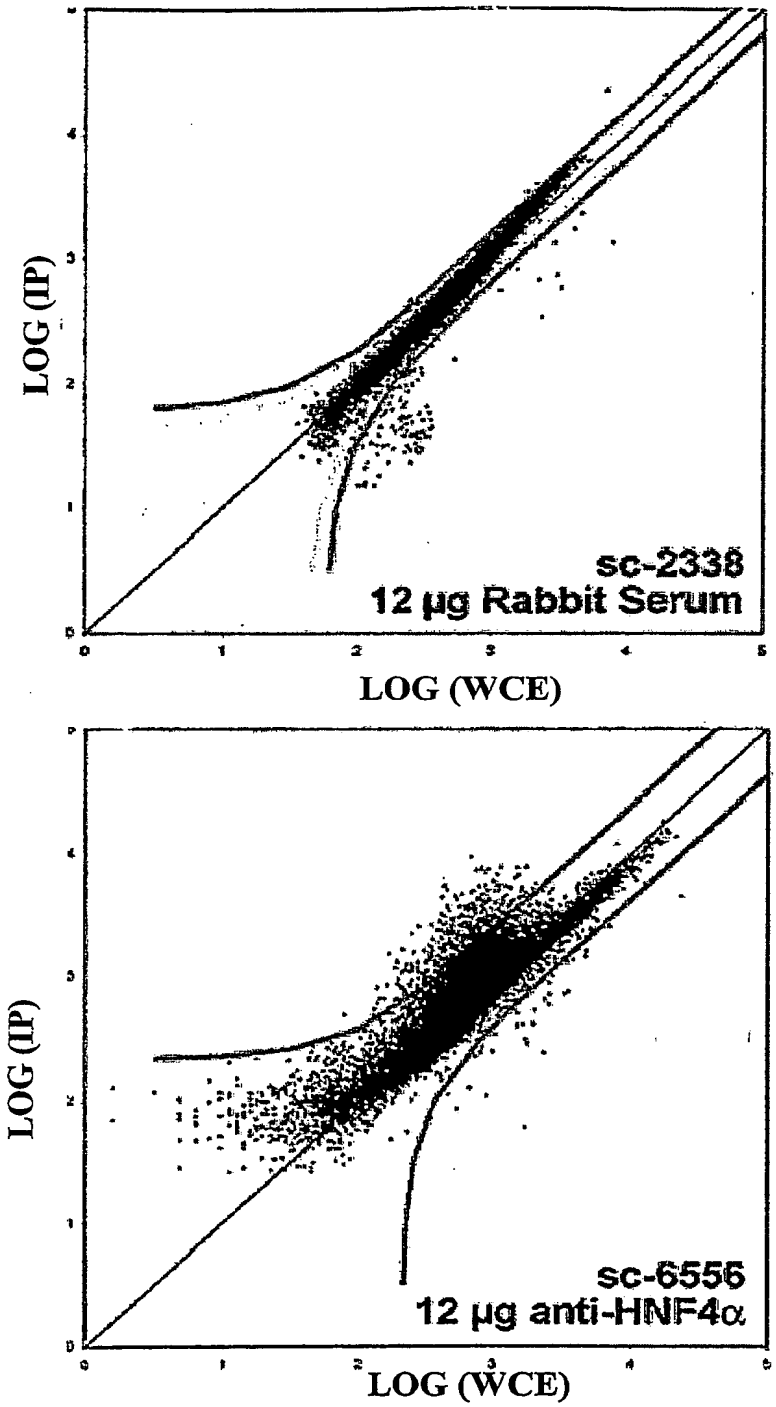

LOG (IP)
LOG (WCE)
sc-2338
12 µg Rabbit Serum
LOG (IP)
LOG (WCE)
sc-6556
12 µg anti-HNF4α

Fig. 8

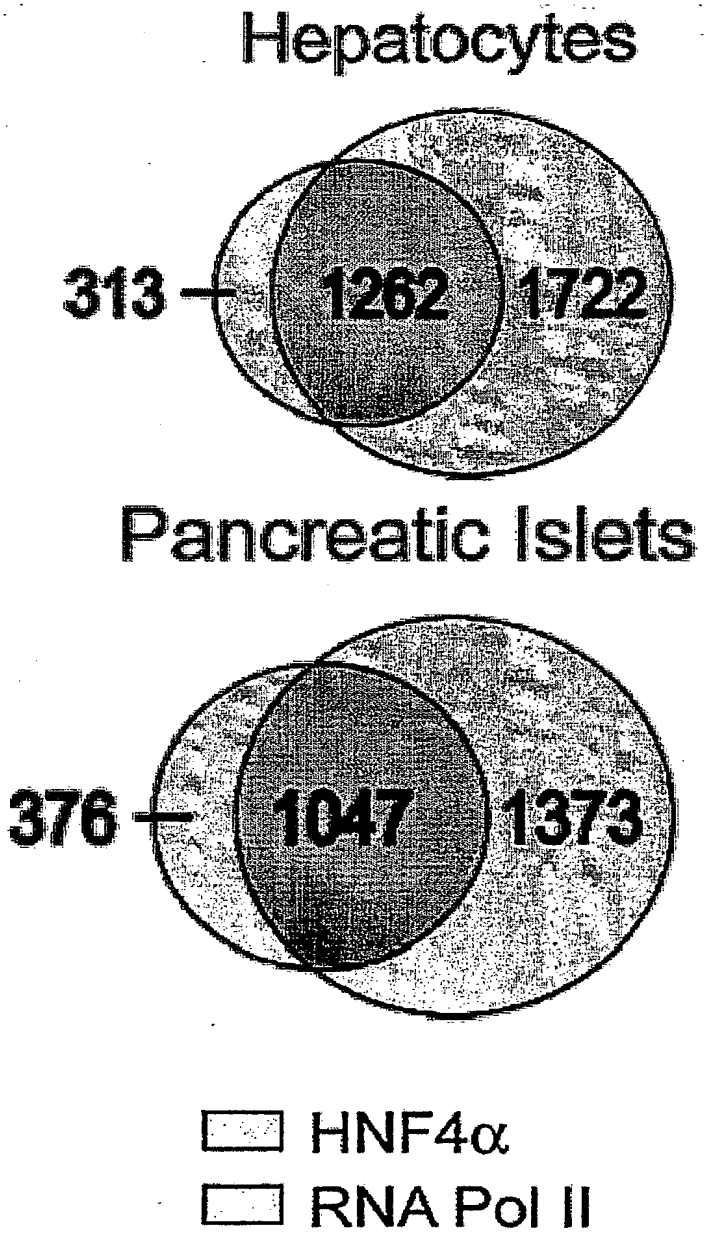

Hepatocytes
313
1262
1722
Pancreatic Islets
376
1047
1373
HNF4α
RNA Pol II

Fig. 9

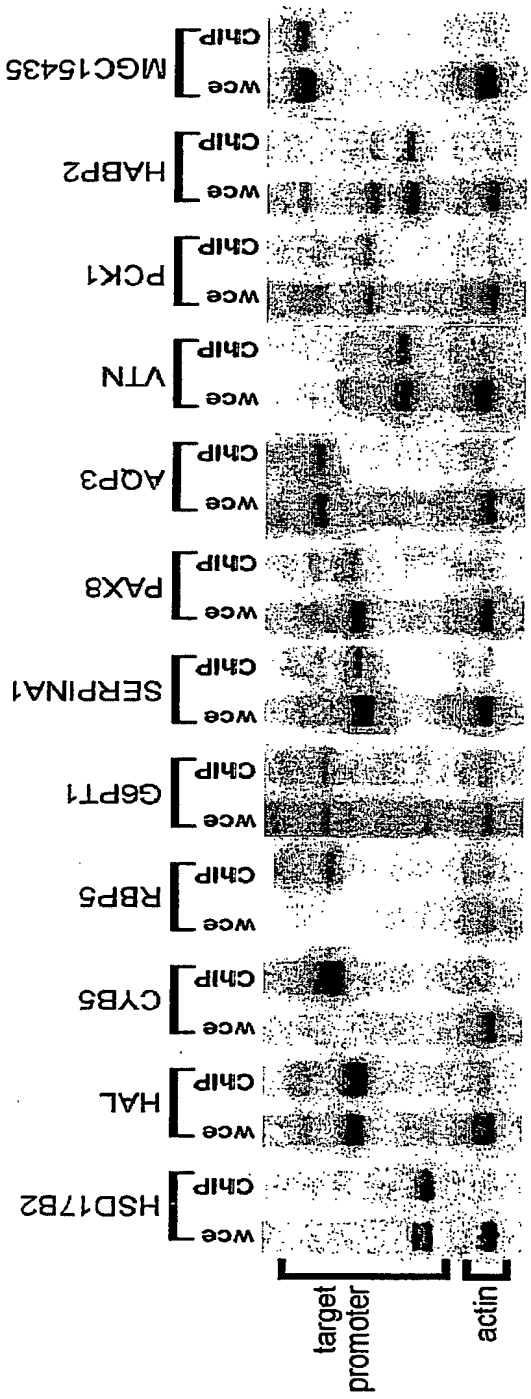

MGC15435
HABP2
PCK1
VTN
AQP3
PAX8
SERPINA1
G6PT1
RBP5
CYB5
HAL
HSD17B2
chip
wce
target promoter
actin

Fig. 10

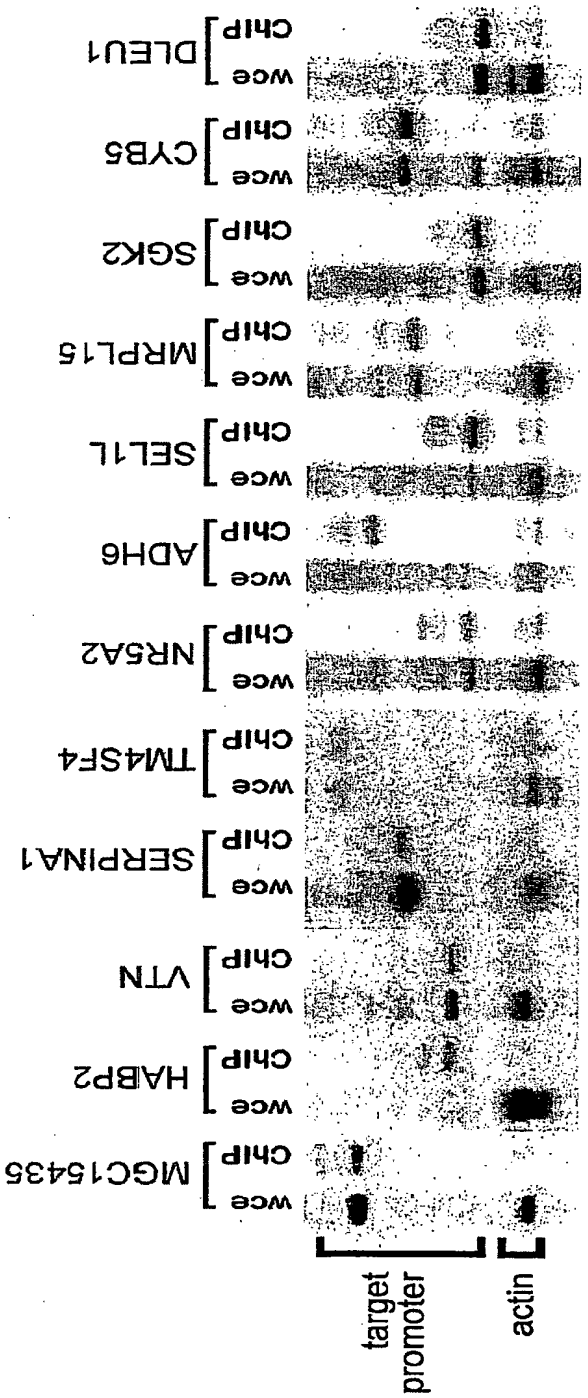

MGC15435
HABP2
VTN
SERPINA1
TM4SF4
NR5A2
ADH6
SEL1L
MRPL15
SGK2
CYB5
DLEU1
wce
ChIP
target promoter
actin

Fig. 11

Name	RefSeq	Description	Hepatocyte	Islets
Chaperone				
C4BPA	NM_000715	complement 4 binding protein a	✓	✓
APCS	NM_001639	amyloid P component	✓	
F11	NM_019559	coagulation factor XI	✓	
C1S	NM_001734	complement component 1s	✓	
VTN	NM_000638	somatomedin B	✓	
Enzyme--Hydrolase				
PGCP	NM_016134	glutamate carboxypeptidase		✓
GLA	NM_000169	galactosidase, alpha		✓
LIPA	NM_000235	lipase A		✓
SPO11	NM_012444	SPO11-like		✓
PAFAH2	NM_000437	platelet-activating factor 2	✓	✓
AADAC	NM_001086	arylacetamide deacetylase	✓	✓
PS-PLA1	NM_015900	phospholipase A1alpha	✓	✓
VNN3	NM_018399	vanin 3	✓	✓
CPB2	NM_016413	carboxypeptidase B2	✓	
ANPEP	NM_001150	alanyl aminopeptidase	✓	
HGFAC	NM_001528	HGF activator	✓	
ENPEP	NM_001977	glutamyl aminopeptidase	✓	
Enzyme--Ligase				
MCCC1	NM_020166	methylcrotonoyl-CoA carboxylase		✓
GARS	NM_002047	glycyl-tRNA synthetase	✓	
TARS	NM_003191	threonyl-tRNA synthetase	✓	
Enzyme--Lyase				
UROD	NM_000374	uroporphyrinogen decarboxylase		✓
PCK1	NM_002591	PEPCK1	✓	
HPCL2	NM_012260	2-hydroxyphytanoyl-CoA lyase	✓	
HAL	NM_002108	histidine ammonia-lyase	✓	
FH	NM_000143	fumarate hydratase	✓	
Enzyme--Oxidoreductase				
COQ7	NM_016138	COQ7 coenzyme Q, 7		✓
ADH4	NM_000670	alcohol dehydrogenase 4		✓
UQCRC2	NM_003366	ubiq.-cyt. c reductase core prot. II	✓	✓
CYB5-M	NM_030579	cytochrome b5	✓	✓
CYP2E	NM_000773	cytochrome P450, IIE	✓	
CYB5	NM_001914	cytochrome b-5	✓	
HSD17B2	NM_002153	hydroxysteroid dehydrogenase 2	✓	
ADH1A	NM_000667	alcohol dehydrogenase 1A	✓	
Enzyme--Transferase				
GCNT3	NM_004751	glucosaminyl transferase 3		✓
FNTB	NM_002028	farnesyltransferase beta	✓	✓
HNMT	NM_006895	histamine N-methyltransferase	✓	
GOT1	NM_002079	aspartate aminotransferase 1	✓	
UGT2B15	NM_001076	UDP glycosyltransferase 2B15	✓	
GBE1	NM_000158	glycogen branching enzyme	✓	
Enzyme Regulator				
SERPING1	NM_000062	C1-inhibitor	✓	
SERPINA1	NM_000295	alpha-1-antitrypsin	✓	
ITIH4	NM_002218	inter-alpha inhibitor H4	✓	
AHSG	NM_001622	alpha-2-HS-glycoprotein	✓	
Ligand Binding				
TMOD2	NM_014548	tropomodulin 2		✓
IGFBP1	NM_000596	IGF binding protein 1	✓	
MT1X	NM_005952	metallothionein 1X	✓	
CRP	NM_000567	C-reactive protein	✓	
APOA2	NM_001643	apolipoprotein A-II	✓	

Name	RefSeq	Description	Hepatocyte	Islets
Signal Transduction--Other				
BIKE	NM_017593	BMP-2 inducible kinase		✓
SGK2	NM_016276	serum/glucocorticoid reg. kinase 2	✓	✓
SEL1L	NM_005065	suppressor of lin-12-like	✓	✓
SCYE1	NM_004757	small cytokine E1	✓	
ANGPTL3	NM_014495	angiopoietin-like 3	✓	
Signal Transduction--Receptor				
HAVCR-1	NM_012206	hepatitis A virus cellular receptor 1		✓
TACR3	NM_001059	tachykinin receptor 3		✓
GNB2L1	NM_006098	GTP binding protein , beta2L1		✓
INSR	NM_000208	insulin receptor		✓
SSTR1	NM_001049	somatostatin receptor 1	✓	✓
TM4SF4	NM_004617	transmembrane 4-4	✓	✓
ASGR2	NM_001181	asialoglycoprotein receptor 2	✓	
GPR39	NM_001508	G protein-coupled receptor 39	✓	
IFNAR1	NM_000629	interferon receptor 1	✓	
TFRC	NM_003234	transferrin receptor	✓	
Transcription Regulation				
ZNF300	NM_052860	kruppel-like zinc finger protein		✓
BCL6	NM_001706	B-cell CLL/lymphoma 6		✓
ZNF155	NM_003445	zinc finger protein 155		✓
FBXO8	NM_012180	F-box only protein 8		✓
NR0B2	NM_021969	Small heterodimer protein	✓	✓
HNF4a7	AF509467	HNF4alpha, alternate splice	✓	✓
NR5A2	NM_003822	LRH-1/FTZ-F1	✓	✓
ELF3	NM_004433	E74-like factor 3	✓	✓
NR1D1	NM_021724	THRA1	✓	
ATF2	NM_001880	activating transcription factor 2	✓	
CREBL2	NM_001310	CREB-like 2	✓	
RARB	NM_016152	RAR-beta	✓	
Transporter--Channel/Pore				
SLC17A2	NM_005835	vesicular glutamate transporter	✓	
AQP3	NM_004925	aquaporin 3	✓	
SLC22A11	NM_018484	hOAT4	✓	
GJB1	NM_000166	gap junction protein, beta 1	✓	
Transporter--Lipids and Small Molecules				
APOH	NM_000042	apolipoprotein H	✓	✓
ALB	NM_000477	albumin	✓	
ABCC2	NM_000392	canalicular OAT	✓	
G6PT1	NM_001467	glucose-6-phosphatase, transport	✓	
Transporter--Proteins				
RAB6KIFL	NM_005733	RAB6 interacting, kinesin-like		✓
PEX13	NM_002618	peroxisome biogenesis factor 13		✓
TMP21	NM_006827	transmembrane trafficking protein		✓
RAB33B	NM_031296	RAS oncogene	✓	✓
NAPA	NM_003827	alpha SNAP	✓	
AP3M1	NM_012095	adaptor-related prot. Complex	✓	
SNX17	NM_014748	sorting nexin 17	✓	

Fig. 12

	BJ-T vs Hepatocytes*		BJ-T vs Pancreatic Islets*	
	BJ-T specific genes	Hepatocyte specific genes	BJ-T specific genes	Islet specific genes
HNF4α/RNA Pol II	19/492 (4%)	996/2389 (42%)	29/546 (5%)	825/1898 (43%)
HNF1α/RNA Pol II	2/492 (.4%)	123/2389 (5.1%)	4/546 (.9%)	32/1898 (1.7%)
HNF6/RNA Pol II	7/492 (1.4%)	105/2389 (4.4%)	3/546 (.5%)	68/1898 (3.6%)

Fig. 13

Name	RefSeq	Name	RefSeq	Name	RefSeq	Name	RefSeq	Name	RefSeq
AADAC	NM_001086	DLEU1	NM_005887	HPX	NM_000613	PHF2	NM_005392	ZNF288	NM_015642
ABCC2	NM_000392	DUSP6	NM_022652	HSD11B1	NM_005525	PIST	NM_020399	ZNF361	NM_018555
ACF	NM_014576	EIF4EBP2	NM_004096	HSD17B2	NM_002153	PLCB1	NM_015192		
ADH1A	NM_000667	ELF3	NM_004433	HSPC111	NM_016391	PLG	NM_000301		
ADH1B	NM_000668	ENPEP	NM_001977	HSPC129	NM_016396	PLGL	NM_002665		
ADH6	NM_000672	F11	NM_019559	IFNAR1	NM_000629	PS-PLA1	NM_015900		
AGT	NM_000029	FE65L2	NM_006051	IGF1R	NM_000875	PZP	NM_002864		
AHSG	NM_001622	FH	NM_000143	IGFBP1	NM_000596	RAB33B	NM_031296		
AK2	NM_001625	FKSG87	NM_032029	INADL	NM_005799	RAMP	NM_016448		
AKR1C2	NM_001354	FLJ10242	NM_018036	ITIH3	NM_002217	RARB	NM_016152		
AKR1C3	NM_003739	FLJ10276	NM_018045	ITIH4	NM_002218	RBP5	NM_031491		
AKR1C4	NM_001818	FLJ10525	NM_018126	ITM2B	NM_021999	RNGTT	NM_003800		
ALB	NM_000477	FLJ10583	NM_018148	KIAA0022	NM_014880	RPL37AP1	NG_000988		
ALDH3A2	NM_000382	FLJ10650	NM_018168	KIAA0669	NM_014779	SAC	NM_018417		
ALS2	NM_020919	FLJ10774	NM_024662	KIAA0844	NM_014951	SCYE1	NM_004757		
AMBP	NM_001633	FLJ11000	NM_018295	KIAA0872	NM_014940	SEL1L	NM_005065		
ANGPTL3	NM_014495	FLJ11838	NM_024664	KIAA1041	NM_014947	SERPINA1	NM_000295		
ANPEP	NM_001150	FLJ12788	NM_022492	KNG	NM_000893	SERPINA10	NM_016186		
AP3M1	NM_012095	FLJ13448	NM_025147	LBP	NM_004139	SERPINA6	NM_001756		
APCS	NM_001639	FLJ13611	NM_024941	LOC51060	NM_015913	SERPINC1	NM_000488		
APG3	NM_022488	FLJ14356	NM_030824	LOC51096	NM_016001	SERPINE1	NM_000602		
APOA2	NM_001643	FLJ20080	NM_017657	LOC51326	NM_016632	SERPING1	NM_000062		
APOH	NM_000042	FLJ20718	NM_017939	LOC54518	NM_019043	SGK2	NM_016276		
AQP3	NM_004925	FLJ21272	NM_025032	LOC56902	NM_020143	SLC17A2	NM_005835		
AQP9	NM_020980	FLJ21934	NM_024743	LOC58486	NM_021211	SLC22A11	NM_018484		
ARHGAP11A	NM_014783	FLJ22551	NM_024708	LY6E	NM_002346	SLPI	NM_003064		
ASGR1	NM_001671	FLJ23259	NM_024727	M17S2	NM_031858	SNX17	NM_014748		
ASGR2	NM_001181	FNTB	NM_002028	M96	NM_007358	SRI	NM_003130		
ATF2	NM_001880	G0S2	NM_015714	MAGEA9	NM_005365	SSA2	NM_004600		
AUTL1	NM_032852	G3A	NM_019101	MGC10500	NM_031477	SSTR1	NM_001049		
BAT3	NM_004639	G6PT1	NM_001467	MGC11034	NM_031453	SSTR4	NM_001052		
BIKE	NM_017593	GARS	NM_002047	MGC11266	NM_024322	STRAIT11499	NM_021242		
BTN2A1	NM_078476	GBE1	NM_000158	MGC13010	NM_032687	SUPV3L1	NM_003171		
C1S	NM_001734	GCKR	NM_001486	MGC15435	NM_032367	SYN3	NM_133632		
C2	NM_000063	GDI2	NM_001494	MGC955	NM_024097	TARS	NM_003191		
C4BPA	NM_000715	GIOT-2	NM_016264	MIA2	NM_054024	TBPL1	NM_004865		
C8B	NM_000066	GJB1	NM_000166	MRPL15	NM_014175	TEF	NM_003216		
CCNE1	NM_001238	GOT1	NM_002079	MRPS18B	NM_014046	TFRC	NM_003234		
CDCA1	NM_031423	GPR39	NM_001508	MSH6	NM_000179	TIEG2	NM_003597		
CISH	NM_013324	GPX2	NM_002083	MT1H	NM_005951	TIEG2	NM_003597		
CLYBL	NM_138280	GRHPR	NM_012203	MT1L	NM_002450	TM4SF4	NM_004617		
CNTNAP2	NM_014141	GTF2B	NM_001514	MT1X	NM_005952	TMEM1	NM_003274		
CPB2	NM_016413	GTF2E1	NM_005513	MTHFD1	NM_005956	TNFRSF6	NM_000043		
CREBL2	NM_001310	GTPBG3	NM_032620	MTP	NM_000253	UGT1A1	NM_000463		
CRP	NM_000567	HABP2	NM_004132	NAPA	NM_003827	UGT2B11	NM_001073		
CTSZ	NM_001336	HAL	NM_002108	NET-2	NM_012338	UGT2B15	NM_001076		
CYB5	NM_001914	HAO1	NM_017545	NFKBIB	NM_002503	UQCRC2	NM_003366		
CYB5-M	NM_030579	HCAP-G	NM_022346	NPC1L1	NM_013389	VNN3	NM_018399		
CYP2E	NM_000773	HGD	NM_000187	NR0B2	NM_021969	VTN	NM_000638		
CYP3A43	NM_022820	HGFAC	NM_001528	NR1D1	NM_021724	WBP4	NM_007187		
DAF	NM_000574	HNF4A	NM_000457	NR5A2	NM_003822	WDF2	NM_052950		
DC13	NM_020188	HNF4A	NM_000457	NRD1	NM_002525	WDR12	NM_018256		
DKFZP564O0463	NM_014156	HNF4a7	AF509467	PAFAH2	NM_000437	XDH	NM_000379		
DKFZP586A0522	NM_014033	HNMT	NM_006895	PAX8	NM_013952	XPC	NM_004628		
DKFZP586M0122	NM_015425	HPCL2	NM_012260	PCK1	NM_002591	ZK1	NM_005815		

Fig. 14

Name	RefSeq
AADAC	NM_001086
ABCC9	NM_020297
ADH4	NM_000670
APOH	NM_000042
ARHGAP11A	NM_014783
B29	NM_031939
BCL6	NM_001706
BIKE	NM_017593
C4BPA	NM_000715
C6orf11	NM_005452
CDC45L	NM_003504
COL3A1	NM_000090
COQ7	NM_016138
CPXCR1	NM_033048
CRH	NM_000756
CTSZ	NM_001336
CYB5-M	NM_030579
DKFZP564J157	NM_018457
DLEU1	NM_005887
DOCK1	NM_001380
DSC1	NM_024421
EIF3S6	NM_001568
ELF3	NM_004433
FBXO8	NM_012180
FE65L2	NM_006051
FIL1(EPSILON)	NM_014440
FLJ10242	NM_018036
FLJ10252	NM_018040
FLJ10474	NM_018104
FLJ10650	NM_018168
FLJ11301	NM_018385
FLJ13273	NM_024751
FLJ13385	NM_024853
FLJ13448	NM_025147
FLJ14855	NM_033210
FLJ20156	NM_017691
FLJ20225	NM_019062
FLJ20234	NM_017720
FLJ20298	NM_017752
FLJ20643	NM_017916
FLJ20731	NM_017946
FLJ21272	NM_025032
FLJ22559	NM_024928
FNTB	NM_002028
GCNT3	NM_004751
GIOT-2	NM_016264
GLA	NM_000169
GNB2L1	NM_006098
GPR74	NM_004885
H4F2	NM_003548
HAVCR-1	NM_012206
HHLA2	NM_007072
HNF4a7	AF509467
IFNA10	NM_002171
INSR	NM_000208
KIAA0101	NM_014736
KIAA0399	NM_015113
KIAA0844	NM_014951
KIF13A	NM_022113
KIR-023GB	NM_015868
KIR2DS2	NM_012312
KIR3DL1	NM_013289
KRTAP1.1	NM_030967
KRTHA3A	NM_004138
LIPA	NM_000235
LOC113201	NM_138423
LOC113220	NM_138424
LOC51092	NM_015996
LOC56906	NM_020147
MCCC1	NM_020166
MGC10500	NM_031477
MGC15677	NM_032878
MIA2	NM_054024
MRPL15	NM_014175
Nod1(-)6kb	NM_006092
NPY2R	NM_000910
NR0B2	NM_021969
NR2C2	NM_003298
NR5A2	NM_003822
PAFAH2	NM_000437
PAX8	NM_013952
pcnp	NM_020357
PEX13	NM_002618
PGCP	NM_016134
PRO2032	NM_018615
PSMA5	NM_002790
PS-PLA1	NM_015900
RAB33B	NM_031296
RAB6KIFL	NM_005733
SDCCAG10	NM_005869
SEL1L	NM_005065
SGK2	NM_016276
SLC26A7	NM_052832
SPO11	NM_012444
SRI	NM_003130
SSTR1	NM_001049
TACR3	NM_001059
TM4SF4	NM_004617
TMOD2	NM_014548
TMP21	NM_006827
UQCRC2	NM_003366
UROD	NM_000374
VNN3	NM_018399
WBP4	NM_007187
ZNF155	NM_003445
ZNF300	NM_052860

Fig. 15A

		Direct	In vitro	Indirect	Sequence Based	ORGANISM
Regulator	Target Gene	Reference	Reference	Reference	Reference	Organism
HNF4α	GST-YA			Paulson 1990		human
HNF4α	TTR		Sladek 1990	Sladek 1990, costa 1991		human
HNF4α	ApoC3		Sladek 1990	Sladek 1990		human
HNF4α	ApoA1		Sladek 1990	Sladek 1990		human
HNF4α	serpina		Sladek 1990	Sladek 1990		human
HNF4α	Pklr		Sladek 1990	Sladek 1990		human
HNF4α	cyp2c13				eguchi 1991	rat
HNF4α	atb		herbst 1991	herbst 1991		rat
HNF4α	ttr		herbst 1991	herbst 1991		rat
HNF4α	hnf1a			tian 1991		human
HNF4α	f9		crossley 1991			human
HNF4α	hnf1a			kuo 1992		human
HNF4α	apob		ladias 1992	ladias 1992		human
HNF4α	ApoC3		ladias 1992	ladias 1992		human
HNF4α	apoa2		ladias 1992	ladias 1992		human
HNF4α	pklr			puzenat 1992		human
HNF4α	f9			reijnen 1992		human
HNF4α	tf			schaeffer 1993		human
HNF4α	hnf1a			zapp 1993		xenopus
HNF4α	pck1		angrand 1994	angrand 1994		rat
HNF4α	pck2		angrand 1994	angrand 1994		rat
HNF4α	cyp2c2		chen 1993	chen 1993		human
HNF4α	cyp2c1		chen 1993	chen 1993		human
HNF4α	cyp2c3		chen 1993	chen 1993		human
HNF4α	cyp7a1		chiang 1994	chiang 1994		rat
HNF4α	ApoA1		fuemkranz 1994	fuemkranz 1994		human
HNF4α	CEACAM1		hauck 1994	hauck 1994		human
HNF4α	apoa4		klistaki 1994	klistaki 1994		human
HNF4α	pklr			liimatta 1994		rat
HNF4α	a2m		matthijs 1994			human
HNF4α	pklr	miquerol 1994				human
HNF4α	rbp2			nakshatri 1994		rodent
HNF4α	otc			nishiyori 1994		mice
HNF4α	acox1		winrow 1994	winrow 1994		rat
HNF4α	hsd17b4		winrow 1994	winrow 1994		rat
HNF4α	f7		erdmann 1995, greenberg 1995	erdmann 1994, greenberg 1995		human
HNF4α	f8		figueiredo 1995	figueiredo 1995		human
HNF4α	epo		galson 1995	galson 1995		human
HNF4α	cyp2c9		ibeanu 1995	ibeanu 1995		human
HNF4α	ambp		rouet 1995	rouet 1995		human
HNF4α	cyp2c23		roussel 1995			rat
HNF4α	cyp2d6		cairns 1996	cairns 1996		human
HNF4α	serpinc1		Fernandez-Rachubinski 1996	Fernandez-Rachubinski 1996		human
HNF4α	bf			garnier 1996		human
HNF4α	f10		hung 1996	hung 1996		human
HNF4α	prlr		moldrup 1996	moldrup 1996		rat
HNF4α	mst1		waltz 1996	waltz 1996		human
HNF4α	lipc			chang 1997		human
HNF4α	g6pc		lin 1997	lin 1997		human
HNF4α	SLC2A2			stoffel 1997		mouse
HNF4α	aldob			stoffel 1997		mouse
HNF4α	gadp			stoffel 1997		mouse
HNF4α	fabp1			stoffel 1997		mouse
HNF4α	cyp2a4		yokomori 1997			mouse
HNF4α	f12		farsetti 1998			human
HNF4α	cyp3a23		huss 1998	huss 1998		rat
HNF4α	shbg		janne 1998	janne 1998		human
HNF4α	apoc2		kardassis 1998	kardassis 1998		human
HNF4α	afp			magee 1998		human
HNF4α	HMGCS2		rodriguez 1998	rodriguez 1998		rodent
HNF4α	ALDH3A1		boesch 1999	boesch 1999		rat
HNF4α	serpina1		hu 1999	hu 1999		human
HNF4α	cyp3a1			ogino 1999		rat
HNF4α	atdh2			pinaire 1999		human
HNF4α	cyp2c12		sasaki 1999	sasaki 1999		rat
HNF4α	GUCY2C		swenson 1999	swenson 1999		human
HNF4α	ang		yanai 1999	yanai 1999		human
HNF4α	ada		dusing 2000			human
HNF4α	hnf6		lahuna 2000	lahuna 2000		human

Fig. 15B

TABLE S4		Direct	In vitro	Indirect	Sequence Based	
Regulator	Target Gene	Reference	Reference	Reference	Reference	Organism
HNF4α	hadhb		nicolas-frances 2000	nicolas-frances 2000		human
[redacted]	pax4		smith 2000	smith 2000		human
[redacted]	ins			wang 2000		mouse
HNF4α	ogdh			wang 2000		mouse
HNF4α	ucp2			wang 2000		mouse
HNF4α	hnf4a		bailly 2001	bailly 2001		human
HNF4α	ghr		jiang 2001	jiang 2001		bovine
HNF4α	cyp3a4			jover 2001		human
HNF4α	cyp3a5			jover 2001		human
HNF4α	cyp3a6			jover 2001		human
HNF4α	cyp2b6			jover 2001		human
HNF4α	cyp2c9			jover 2001		human
HNF4α	fmo1			huo 2001		rabbit
HNF4α	cyp3a16		nakayama2001	nakayama2001		mouse
HNF4α	akr1c4		ozeki 2001	ozeki 2001		human
HNF4α	cyp8b1		zhang 2001	zhang 2001		human
HNF4α	hpd		aarenstrup 2002	aarenstrup 2002		rat
HNF4α	cyp27		garuti 2002	garuti 2002		human
HNF4α	NOS2A		guo 2002	guo 2002		rat
HNF4α	cpt1a			louet 2002		human
HNF4α	ppara		pineda-torra 2002	pineda-torra 2002		human
HNF4α	gk		roth 2002			rat
HNF4α	Serpina1	Soutoglou 2002				human
HNF1α	FGA			baumhueter 1990		
HNF1α	FGB			baumhueter 1990		
HNF1α	FGG			baumhueter 1990		
HNF1α	afp			baumhueter 1990		
HNF1α	serpina1			baumhueter 1990		
HNF1α	afm			herbst 1991	cereghini 1990	rat (herb
HNF1α	afm			tronche 1991		rat
HNF1α	cyp2e1		gonzalez 1990, hayashi 1991			animal
HNF1α	aldob		raymondjean 1991			rat
HNF1α	aldob		ito 1990			rat
HNF1α	igfbp1			suwanichkul 1990, babajko 1993		human
HNF1α	igfbp1			powell 1993		human
HNF1α	igfbp1			suh 1995, suh 1997		rat
HNF1α	crp			toniatti 1990		
HNF1α	apoa2			chambaz 1991		human
HNF1α	ttr			costa 1991		mouse
HNF1α	ttr			herbst 1991		rat
HNF1α	hdlbp				drewes 1991	xenopus
HNF1α	rbp5			tripodi 1991		human
HNF1α	f2		bancroft 1992	bancroft 1992		human
HNF1α	apob		brooks 1992			human
HNF1α	insr		cameron 1992			human
HNF1α	insr		cameron 1992			human
HNF1α	agt			congiu 1992		mouse
HNF1α	ins			emens 1992		rat
HNF1α	pklr		puzenat 1992			
HNF1α	tat		schweizer-groyer 1992			rat
HNF1α	siat1		svensson 1992	svensson 1992, bois-joyeux 1995		human
HNF1α	adh1			van ooji 1992		human
HNF1α	crhbp				behan 1993	human
HNF1α	afp			bernier 1993		human
HNF1α	fgb		dalmon 1993	dalmon 1993		human
HNF1α	lyz				grajer 1993	chicken
HNF1α	aldob			gregori 1993		
HNF1α	fbg			hayashi 1993		human
HNF1α	apoa1			krilis 1993		
HNF1α	apoc3			krilis 1993		
HNF1α	crp		li 1996	ku 1993, li 1996		mouse
HNF1α	fgb			roberts 1993		xenopus
HNF1α	proc			berg 1994		human
HNF1α	serpina1			bulla 1994		
HNF1α	gsta2		clairmont 1994			human
HNF1α	cyp2c13			legraverend 1994		human
HNF1α	pklr	miqueral 1994				human
HNF1α	anpep		olsen 1994	olsen 1994		human
HNF1α	si		wu 1994	wu 1994		human

Fig. 15C

TABLE S4		Direct	In vitro	Indirect	Sequence Based	
	Target Gene	Reference	Reference	Reference	Reference	Organism
HNF1α	C4BPA			arenzana 1995		human
HNF1α	FGA			hu 1995		human
HNF1α	igf1			kufik 1995, nolten 1995		salmon, human
HNF1α	cyp2e1		liu 1995	liu 1995, lerche 1996		rat
HNF1α	ambp		rouet 1995	rouet 1995		human
HNF1α	ddc		aguanno 1996	aguanno 1996		human
HNF1α	f8		mcglynn 1996	mcglynn 1996		human
HNF1α	plg		meroni 1996	meroni 1996		human
HNF1α	pah			pontoglio 1996		mouse
HNF1α	hmgcs2				boukaftane 1997	human
HNF1α	lipc			chang 1997		rat
HNF1α	cyp2h1		dogra 1997	dogra 1997		chicken
HNF1α	ugt2b1		hansen 1997	hansen 1997		human, rat
HNF1α	guanylin		hochman 1997	hochman 1997		mouse
HNF1α	g6p		lin 1997	lin 1997		human
HNF1α	cyp2e1		McGehee 1997	McGehee 1997		rodent
HNF1α	pah		Pontoglio 1997			mouse
HNF1α	ipaI		Taylor 1997			mouse
HNF1α	hnf4a			bailly 1998		rat
HNF1α	hnf3a			bailly 1998		rat
HNF1α	cebpa			bailly 1998		rat
HNF1α	g6pc		lin 1999	lin 1998		human
HNF1α	afp		magee 1998	magee 1998		human
HNF1α	SLC5A1		rhoads 1998			rat
HNF1α	si			rodolosse 1998		human
HNF1α	gc		song y 1998	song y 1998		human
HNF1α	SULT2A1		song c 1998	song c 1998		rat
HNF1α	proc			spek 1998		human
HNF1α	g6pc		streeper 1998	streeper 1998		human
HNF1α	SLC10A1		trauner 1998			human
HNF1α	Ins			wang 1998		human
HNF1α	ugt1a1		bernard 1999			human, mouse
HNF1α	cyp7a1		chen 1999			human
HNF1α	dpp6			erickson 1999		human
HNF1α	serpina6		hu 1999	hu 1999		human
HNF1α	igf1			meton 1999		salmon
HNF1α	ins		okita 1999	okita 1999		human
HNF1α	CYP27A1		rao 1999	rao 1999		rat
HNF1α	lct		spodsberg 1999	spodsberg 1999		mice
HNF1α	SLC5A1			wood 1999		human
HNF1α	fabp1			akiyama 2000		mouse
HNF1α	cyp7a1		antes 2000	antes 2000		mice
HNF1α	slc2a2		cha 2000	cha 2000		human
HNF1α	dpp6		erickson 2000	erickson 2000		human
HNF1α	UGT2B17		gregory 2000	gregory 2000		human
HNF1α	UGT2B7		ishii 2000	ishii 2000		human
HNF1α	ugt1a7		metz 2000	metz 2000		rat
HNF1α	fech			muppala 2000		mouse
HNF1α	gjb1		piechocki 2000	piechocki 2000		human
HNF1α	SLC5A2		Pontoglio 2000	pontoglio 2000		human
HNF1α	pax4		smith 2000	smith 2000		human
HNF1α	ogdh			wang 2000		rat
HNF1α	aldob			wang 2000		rat
HNF1α	ins			wang 2000		rat
HNF1α	SLC5A2			wang 2000		rat
HNF1α	pklr			wang 2000		rat
HNF1α	hmgcr			wang 2000		rat
HNF1α	hnf4a		bailly 2001	bailly 2001		human
HNF1α	pdx1		ben-shushan 2001	ben-shushan 2001		human
HNF1α	hnf4a7	Boj 2001				mouse
HNF1α	hnf3g	Boj 2001				mouse
HNF1α	hnf4g	Boj 2001				mouse
HNF1α	gck		cha 2001	cha 2001		human
HNF1α	hnf4a	Hatzis 2001	hatzis 2001	hatzis 2001		human
HNF1α	g6pc			hiraiwa 2001		mouse
HNF1α	g6pt1			hiraiwa 2001		mouse
HNF1α	slc21a6		jung 2001	jung 2001		human
HNF1α	slc21a8			jung 2001		human
HNF1α	ngn3			lee 2001		human
HNF1α	igfbp1			leu 2001		rodent
HNF1α	g6p			leu 2001		rodent
HNF1α	afp			leu 2001		rodent
HNF1α	fmo1		luo 2001	luo 2001		rabbit, human

Fig. 15D

TABLE S4		Direct	In vitro	Indirect	Sequence Based	
Regulator	Target Gene	Reference	Reference	Reference	Reference	Organism
HNF1α	CYP27A1		memom 2001			hamster
HNF1α	AKR1C4		oxeki 2001	ozeki 2001		human
HNF1α	NR5A2		pare 2001	pare 2001		mouse
HNF1α	cyp2c11		park 2001	park 2001		rodent
HNF1α	cyp2a2		park 2001	park 2001		rodent
HNF1α	cyp4a2		park 2001	park 2001		rodent
HNF1α	pklr		parrizas 2001			human
HNF1α	slc2a2		parrizas 2001			human
HNF1α	pah		parrizas 2001			human
HNF1α	c8a			pontoglio 2001		mouse
HNF1α	c5			pontoglio 2001		mouse
HNF1α	cyp2e1		roe 2001			rat
HNF1α	nr1h4		shih 2001	shih 2001		mouse
HNF1α	SLC10A2		shih 2001	shih 2001		mouse
HNF1α	SLC17A1			soumounou 2001		human, mouse
HNF1α	hnf4a7			thomas 2001		human
HNF1α	ins			yamakawa 2001		human
HNF1α	Nr5a2			zhang 2001		
HNF1α	SLC5A1			vayro 2001		sheep
HNF1α	slc2a2		ban 2002	ban 2002		human
HNF1α	si			boudreau 2002		mouse
HNF1α	SLC17A1			cheret 2002		mouse
HNF1α	SLC10A1		geier 2002			rat
HNF1α	UGT2B17		gregory 2002	gregory 2002		human
HNF1α	hnf4a7			hansen 2002		mouse
HNF1α	gjb1			koffler 2002		rat
HNF1α	AKR1C4		ozeki 2002	ozeki 2002		human
HNF1α	cldn2			sakaguchi 2002		human, mouse
HNF1α	fgfr4		shah 2002	shah 2002		human
HNF1α	igf1			yang 2002		human/rat
HNF1α	mif			yang 2002		human/rat
HNF1α	Serpina1	Soutoglou 2002				human
HNF1α	c1		zahedi 2002			human

Fig. 16

Name	RefSeq	Name	RefSeq	Name	RefSeq	Name	RefSeq	Name	RefSeq
A1BG	NM_130786	DHFR	NM_000791	GSS	NM_000178	ORC1L	NM_004153	UGT1A1	NM_000463
AASS	NM_005763	DKFZP434J037	NM_030952	H3FF	NM_003533	PABPC1	NM_002568	UGT2B11	NM_001073
ABCA8	NM_007168	DKFZP564O0523	NM_032120	H4FK	NM_003546	PCDHA12	NM_018903	UGT2B15	NM_001076
ABCB11	NM_003742	DKFZP586A0522	NM_014033	HABP2	NM_004132	PCK1	NM_002591	URKL1	NM_017859
ABCC2	NM_000392	DXF68S1E	NM_012080	HBP1	NM_012257	PHTF1	NM_006608	VCP	NM_007126
ABL2	NM_007314	E2F1	NM_005225	HCAP-G	NM_022346	PIK4CB	NM_002651	VTN	NM_000638
ACVR1	NM_001105	E2F1	NM_005225	HESX1	NM_003865	PLGL	NM_002665	WDR12	NM_018256
ADH1A	NM_000667	EIF4A1	NM_001416	HIVEP3	NM_024503	POLR2D	NM_004805	WDR5B	NM_019069
ADH1B	NM_000668	EIF4E	NM_001968	HMGCR	NM_000859	POLS	NM_006999		
AF038169	NM_013310	ELOVL1	NM_016031	HNF4a7	AF509467	PON1	NM_000446		
AGTR1	NM_000685	EPHA1	NM_005232	HNMT	NM_006895	PPFIA1	NM_003626		
AKR1C4	NM_001818	F11	NM_019559	HNRPR	NM_005826	PPP2R5A	NM_006243		
ALDH3A1	NM_000691	F9	NM_000133	HSD17B4	NM_000414	PRO1855	NM_018509		
ALDH5A1	NM_001080	FABP5	NM_001444	HSP105B	NM_006644	PSMA1	NM_002786		
AMBP	NM_001633	FACTP140	NM_007192	HSPA1B	NM_005346	PSMB1	NM_002793		
AMT	NM_000481	FADS3	NM_021727	HTR2B	NM_000867	PTPRR	NM_002849		
APCS	NM_001639	FLJ10209	NM_018026	IF	NM_000204	REA	NM_007273		
APOH	NM_000042	FLJ10407	NM_018087	INSM2	NM_032594	RING1	NM_002931		
ASPA	NM_000049	FLJ10415	NM_018089	IRF3	NM_001571	RNF20	NM_019592		
BCAR1	NM_014567	FLJ10578	NM_018144	IRF6	NM_006147	RPL35	NM_007209		
BCKDHA	NM_000709	FLJ10650	NM_018168	ITGAV	NM_002210	RPL37AP1	NG_000988		
BF	NM_001710	FLJ11029	NM_018304	ITIH1	NM_002215	RPLP1	NM_001003		
BM039	NM_018455	FLJ11105	NM_018323	JIK	NM_016281	RPS6KA5	NM_004755		
BNIP3L	NM_004331	FLJ11301	NM_018385	KIAA0806	NM_014813	RRP46	NM_020158		
BTN3A2	NM_007047	FLJ11726	NM_024971	KIAA0872	NM_014940	SART3	NM_014706		
C1S	NM_001734	FLJ11773	NM_021934	KIAA1056	NM_014894	SAS10	NM_020368		
C2	NM_000063	FLJ12552	NM_022832	KLF3	NM_016531	SCYB13	NM_006419		
C20orf188	NM_015638	FLJ12770	NM_032174	LIMK1	NM_016735	SEC10L1	NM_006544		
C8B	NM_000066	FLJ12910	NM_024573	LOC51060	NM_015913	SERPING1	NM_000062		
C8G	NM_000606	FLJ13798	NM_024773	LOC51074	NM_015957	SERPINI1	NM_005025		
CACNA1D	NM_000720	FLJ14153	NM_022736	LOC51287	NM_016565	SILV	NM_006928		
CASP2	NM_032982	FLJ20084	NM_017659	LOC51633	NM_016023	SLC1A3	NM_004172		
CCT8	NM_006585	FLJ20156	NM_017691	LOC51646	NM_016061	SLC25A13	NM_014251		
CDC25A	NM_001789	FLJ20422	NM_017814	LOC56906	NM_020147	SLC7A9	NM_014270		
CDC2L5	NM_003718	FLJ20627	NM_017909	LOC81558	NM_030802	SMARCC1	NM_003074		
CDK2	NM_001798	FLJ20671	NM_017924	LOH11CR2A	NM_014622	SMCY	NM_004653		
CDSN	NM_001264	FLJ20772	NM_017956	M17S2	NM_031858	SNRPD2	NM_004597		
CFL1	NM_005507	FLJ21934	NM_024743	MAP2K5	NM_002757	SNW1	NM_012245		
CH25H	NM_003956	FLJ21963	NM_024560	MGC10500	NM_031477	SNX3	NM_003795		
CLCN3	NM_001829	FLJ22169	NM_024085	MGC13053	NM_032710	SPG4	NM_014946		
CLDN2	NM_020384	FLJ22557	NM_024713	MGC16169	NM_033115	SPINK1	NM_003122		
CLLD8	NM_031915	FLJ23071	NM_025192	MGC16386	NM_080668	SPP2	NM_006944		
COL5A1	NM_000093	FLJ23263	NM_025115	MGC4189	NM_032308	SRF	NM_003131		
COL5A3	NM_015719	FLJ23375	NM_024956	MGST3	NM_004528	STMN2	NM_007029		
COPB2	NM_004766	FLJ23499	NM_022761	MN1	NM_002430	TAF2GL	NG_001012		
COPS7A	NM_016319	FLJ23598	NM_024783	NEK6	NM_014397	TAT	NM_000353		
CRADD	NM_003805	FXYD7	NM_022006	NFKBIA	NM_020529	TBX2	NM_005994		
CRI1	NM_014335	G6PC	NM_000151	NFKBIA	NM_020529	TCEB3	NM_003198		
CRP	NM_000567	GABPA	NM_002040	NFKBIA	NM_020529	TM4SF4	NM_004617		
CSN2	NM_001891	GAL3ST2	NM_033036	NOLC1	NM_004741	TMF1	NM_007114		
CYGB	NM_134268	GBF1	NM_004193	NR1I2	NM_022002	TMOD2	NM_014548		
CYP3A43	NM_022820	GJB1	NM_000166	NTF2	NM_005796	TNFRSF6	NM_000043		
CYP51	NM_000786	GRB2	NM_002086	OAT	NM_000274	TNFSF10	NM_003810		
D13S106E	NM_005800	GRO1	NM_001511	OAZ2	NM_002537	TOMM70A	NM_014820		
DDB2	NM_000107	GRO3	NM_002090	OGFR	NM_007346	TSG101	NM_006292		

Fig. 17

Name	RefSeq	Name	RefSeq	Name	RefSeq	Name	RefSeq
AASS	NM_005763	FLJ11271	NM_018373	JIK	NM_016281	SEMA6A	NM_020796
ABCB8	NM_007188	FLJ11301	NM_018385	KIAA0660	NM_012297	SERPINB8	NM_002640
ACPP	NM_001099	FLJ11773	NM_021934	KIAA0712	NM_014715	SERPING1	NM_000062
ACVR1	NM_001105	FLJ12770	NM_032174	KIAA0806	NM_014813	SERPINI1	NM_005025
ADH1A	NM_000667	FLJ12910	NM_024573	KIAA0872	NM_014940	SH3BGRL	NM_003022
AF038169	NM_013310	FLJ13220	NM_021927	KIAA1056	NM_014894	SLC1A3	NM_004172
AF15Q14	NM_020380	FLJ13798	NM_024773	KRTAP1.1	NM_030967	SNRPD2	NM_004597
AGT	NM_000029	FLJ13955	NM_024759	LAMC2	NM_018891	SNW1	NM_012245
AMBP	NM_001633	FLJ14153	NM_022736	LBC	NM_006738	SPG4	NM_014946
AMT	NM_000481	FLJ14486	NM_032792	LOC51060	NM_015913	SPINK1	NM_003122
APCS	NM_001639	FLJ20084	NM_017659	LOC51287	NM_016565	TEGT	NM_003217
APOH	NM_000042	FLJ20156	NM_017691	LOC51633	NM_016023	TMF1	NM_007114
ARL1	NM_001177	FLJ20422	NM_017814	LOC56906	NM_020147	TNFRSF6	NM_000043
BBP	NM_032027	FLJ20627	NM_017909	LOC81558	NM_030802	TNFRSF6	NM_000043
BCKDHA	NM_000709	FLJ20643	NM_017916	LOH11CR2A	NM_014622	TNFRSF6	NM_000043
BF	NM_001710	FLJ20671	NM_017924	LUC7A	NM_016424	TNFRSF6	NM_000043
BTN3A2	NM_007047	FLJ20772	NM_017956	MDH1	NM_005917	TNFSF10	NM_003810
C1S	NM_001734	FLJ21272	NM_025032	MDS029	NM_018464	TOMM70A	NM_014820
C20orf188	NM_015638	FLJ21934	NM_024743	MEIS1	NM_002398	UGT2B15	NM_001076
C2F	NM_006331	FLJ21963	NM_024560	MGC13040	NM_032930	UGT2B17	NM_001077
C8orf4	NM_020130	FLJ22169	NM_024085	MGC13053	NM_032710	VCP	NM_007126
CCT8	NM_006585	FLJ23263	NM_025115	MGC19595	NM_033415	VTN	NM_000638
CDC2L5	NM_003718	FLJ23375	NM_024956	MGC3020	NM_024048	WDR12	NM_018256
CH25H	NM_003956	GABARAPL1	NM_031412	MGC3413	NM_032678	ZNF317	NM_020933
CIR	NM_004882	GABPA	NM_002040	MGC4189	NM_032308		
CLCN4	NM_001830	GCP3	NM_006322	MGST3	NM_004528		
CLDN2	NM_020384	GJB1	NM_000166	MTERF	NM_006980		
CLLD8	NM_031915	GLA	NM_000169	NET-6	NM_014399		
CLNS1A	NM_001293	GRB2	NM_002086	NOLC1	NM_004741		
CLONE24922	NM_015679	GRO1	NM_001511	NOVA1	NM_006489		
CMG1	NM_025103	GRO3	NM_002090	NR0B2	NM_021969		
COPB2	NM_004766	GSS	NM_000178	NUDT2	NM_001161		
COPS7A	NM_016319	GSTA4	NM_001512	OGFR	NM_007346		
COX4I1	NM_001861	GTF2E1	NM_005513	ORC1L	NM_004153		
COX7A2L	NM_004718	H4FA	NM_003538	PAPA-1	NM_031288		
CRI1	NM_014335	H4FH	NM_003543	PEX6	NM_000287		
CSN2	NM_001891	HABP2	NM_004132	PMAIP1	NM_021127		
CYP3A43	NM_022820	HASJ4442	NM_017528	PPFIA1	NM_003626		
DKFZp761D221	NM_032291	HBOA	NM_007067	PPFIBP1	NM_003622		
DKFZp761J139	NM_032280	HBP1	NM_012257	PPP1R3D	NM_006242		
EED	NM_003797	HLA-G	NM_002127	PSMA1	NM_002786		
EGR2	NM_000399	HMG2	NM_002129	PSMB1	NM_002793		
EHD4	NM_014599	HNF4a7	AF509467	PTPRN2	NM_002847		
EHF	NM_012153	HNRPA2B1	NM_031243	REA	NM_007273		
EIF4E	NM_001968	HNRPR	NM_005826	RECK	NM_021111		
F11	NM_019559	HSD17B4	NM_000414	RIG-I	NM_014314		
F2RL2	NM_004101	HSN44A4A	NM_015372	RPC32	NM_006467		
FABP5	NM_001444	HSP105B	NM_006644	RPL36P1	NG_000983		
FER1L3	NM_133337	HSPA1B	NM_005346	RPS6KA5	NM_004755		
FLJ10342	NM_018064	HSPC125	NM_014165	RRP46	NM_020158		
FLJ10407	NM_018087	HT007	NM_018480	SAMHD1	NM_015474		
FLJ10415	NM_018089	HTR2B	NM_000867	SART3	NM_014706		
FLJ10482	NM_018107	humNRDR	NM_021004	SAS10	NM_020368		
FLJ10650	NM_018168	IGSF3	NM_001542	SCYA28	NM_019846		
FLJ11029	NM_018304	IRF3	NM_001571	SEC10L1	NM_006544		

Fig. 18A

Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq
24432	NM_022914	APOA5	NM_052968	C3F	NM_005768	CPT1B	NM_004377	DNAJA3	NM_005147	FLJ11184	NM_018352	FLJ22028	NM_024854
384D8-2	NM_014551	APOB	NM_000384	C40	NM_017546	CPT2	NM_000098	DNAJB11	NM_016306	FLJ11186	NM_018353	FLJ22169	NM_024085
54TM	NM_020470	APOC2	NM_000483	C4A	NM_007293	CRADD	NM_003805	DNAJB4	NM_007034	FLJ11198	NM_018358	FLJ22191	NM_025231
A1BG	NM_130786	APOC3	NM_000040	C4B	NM_000592	CREBL2	NM_001310	DNAJB9	NM_012328	FLJ11274	NM_018375	FLJ22353	NM_024587
AASS	NM_005763	APOH	NM_000042	C4BPA	NM_000715	CRFG	NM_012341	DOC-1R	NM_005851	FLJ11286	NM_018381	FLJ22477	NM_024735
AB026190	NM_014458	AQP3	NM_004925	C6orf11	NM_005452	CRI1	NM_014335	DPAGT1	NM_001382	FLJ11342	NM_018394	FLJ22501	NM_024747
ABCA6	NM_080284	AQP6	NM_001652	C6orf35	NM_018452	CRIPT	NM_014171	DPM1	NM_003859	FLJ11526	NM_024632	FLJ22551	NM_024708
ABCB10	NM_012089	AQP9	NM_020980	C7orf10	NM_024728	CRKL	NM_005207	DSCR3	NM_006052	FLJ11728	NM_024971	FLJ22555	NM_024520
ABCB11	NM_003742	ARF1GAP	NM_018209	C8B	NM_000066	CROT	NM_021151	DUSP11	NM_003584	FLJ11767	NM_024593	FLJ22557	NM_024713
ABCC2	NM_000392	ARFD1	NM_001656	C8G	NM_000606	CRP	NM_000567	DUSP3	NM_004090	FLJ11838	NM_024664	FLJ22578	NM_024864
ABCC3	NM_003786	ARG2	NM_001172	C8orf4	NM_020130	CRSP3	NM_004830	DUSP6	NM_022652	FLJ11848	NM_025155	FLJ22637	NM_025165
ABCC6	NM_001171	ARHGAP11A	NM_014783	CABC1	NM_020247	CRSP6	NM_004270	DYRK1B	NM_004714	FLJ12171	NM_024619	FLJ22649	NM_021928
ABCE1	NM_002940	ARHI	NM_004675	CACNA2D2	NM_006030	CRY1	NM_004075	EEF1B2	NM_021121	FLJ12377	NM_024989	FLJ22692	NM_025049
ABCG1	NM_004915	ARL1	NM_001177	CACNA2D2	NM_006030	CRYZ	NM_001889	EFG1	NM_024996	FLJ12439	NM_023077	FLJ22729	NM_024683
ABCG8	NM_022437	ARL5	NM_012097	CAMK2D	NM_001221	CS	NM_004077	EHD3	NM_014600	FLJ12552	NM_022832	FLJ22865	NM_025109
ABLIM	NM_006720	ARL7	NM_005737	CARD15	NM_022162	CSDUFD1	NM_031919	EHHADH	NM_001966	FLJ12618	NM_024884	FLJ22875	NM_032231
ABS	NM_016222	ARPC5	NM_005717	CASP2	NM_032982	CSNK2A1	NM_001895	EHM2	NM_019114	FLJ12707	NM_022067	FLJ23071	NM_025192
ABT1	NM_013375	ARS2	NM_015908	CASP6	NM_001226	CSPG6	NM_005445	EIF2S1	NM_004094	FLJ12770	NM_032174	FLJ23093	NM_024643
ACAA2	NM_006111	ASB3	NM_016115	CAT56	NM_025263	CSTF1	NM_001324	EIF2S3	NM_001415	FLJ12788	NM_022492	FLJ23109	NM_024814
ACADSB	NM_001609	ASGR1	NM_001671	CATSPER	NM_053054	CSTF3	NM_001326	EIF4E	NM_001968	FLJ12886	NM_019106	FLJ23251	NM_024818
ACADVL	NM_000018	ASGR2	NM_001181	CBARA1	NM_006077	CTMP	NM_053055	EIF4EBP2	NM_004096	FLJ12889	NM_024945	FLJ23263	NM_025115
ACF	NM_014576	ATF2	NM_001880	CBS	NM_000071	CTSZ	NM_001336	EIF5	NM_001969	FLJ12910	NM_024573	FLJ23305	NM_025059
ACLY	NM_001096	ATF4	NM_001675	CBX3	NM_007276	CUL2	NM_003591	ELF3	NM_004433	FLJ13102	NM_024887	FLJ23441	NM_024678
ACO2	NM_001098	ATF7	NM_006856	CBX5	NM_012117	CYB5	NM_001914	ELP2	NM_018255	FLJ13158	NM_024909	FLJ23468	NM_024629
ACOX1	NM_004035	ATM	NM_000051	CCNG1	NM_004060	CYB5-M	NM_030579	ENC1	NM_003633	FLJ13162	NM_025002	FLJ23499	NM_022761
ACOX3	NM_003501	ATP5C1	NM_005174	CCNG2	NM_004354	CYP1A2	NM_000761	EPB72	NM_004099	FLJ13181	NM_025188	FLJ23516	NM_024725
ACP2	NM_001610	ATP5F1	NM_001688	CCNH	NM_001239	CYP1B1	NM_000104	EPHA2	NM_004431	FLJ13194	NM_025148	FLOT1	NM_005803
ACTA2	NM_001613	ATP5G3	NM_001689	CCT6A	NM_001762	CYP21A2	NM_000500	EPI64	NM_031937	FLJ13195	NM_022906	FMR1	NM_002024
ACTN1	NM_001102	ATP6D	NM_004691	CD1D	NM_001766	CYP2B6	NM_000767	ERBB2IP	NM_018695	FLJ13262	NM_024914	FNTA	NM_002027
ACTR3	NM_005721	ATP6G1	NM_004888	CD68	NM_001251	CYP2C8	NM_000770	ERBB3	NM_001982	FLJ13273	NM_024751	FNTB	NM_002028
ACVR1	NM_001105	ATP6L	NM_001694	CDA	NM_001785	CYP2D6	NM_000106	ERCC5	NM_000123	FLJ13291	NM_032178	FOSL2	NM_005253
ACY1	NM_000666	ATP6M	NM_015994	CDC14A	NM_003672	CYP2D7AP	NG_000853	ERCC6	NM_000124	FLJ13340	NM_025085	FRG1	NM_004477
ADD2	NM_016814	ATP6S14	NM_004231	CDC25A	NM_001789	CYP2E	NM_000773	ERO1L	NM_014584	FLJ13448	NM_025147	FRK	NM_002031
ADD34	NM_031480	ATP7B	NM_000053	CDC42BPB	NM_006035	CYP2J2	NM_000775	EVA1	NM_005797	FLJ13491	NM_024623	FSTL3	NM_005860
AD158	NM_032270	ATPaW	NM_015684	CDC5L	NM_001253	CYP3A43	NM_022820	EVC	NM_014556	FLJ13611	NM_024941	FTHFD	NM_012190
AD24	NM_022451	AUP1	NM_012103	CDCA1	NM_031423	CYP3A5	NM_000777	EVG1	NM_032561	FLJ13615	NM_025114	FTSJ1	NM_012280
ADH1B	NM_000668	AUTL1	NM_032852	CDK2	NM_001798	CYP4F11	NM_021187	EWSR1	NM_013986	FLJ13660	NM_025197	FUBP1	NM_003902
ADH6	NM_000672	B29	NM_031939	CDKL3	NM_016508	CYP4F2	NM_001082	F10	NM_000504	FLJ13769	NM_025012	FXYD7	NM_022006
ADPRH	NM_001125	B3GAT1	NM_018644	CDKN1B	NM_004064	CYP4F3	NM_000896	F12	NM_000505	FLJ13798	NM_024773	FZD1	NM_003505
ADPRTL1	NM_006437	b58b5R	NM_016230	CDKN1B	NM_004064	CYP51	NM_000786	F7	NM_019616	FLJ13949	NM_025077	FZD3	NM_017412
ADPRTL3	NM_005485	BACE	NM_012104	CDKN1B	NM_004064	CYP6B1	NM_004391	F9	NM_000133	FLJ13952	NM_024798	GDS2	NM_015714
ADRB2	NM_000024	BAI2	NM_001703	CDKN1B	NM_004064	Cyt19	NM_020582	FACTP140	NM_007192	FLJ13962	NM_024862	G10	NM_003910
AF093680	NM_013242	BAL	NM_031458	CDSN	NM_001264	D123	NM_000023	FAP48	NM_053274	FLJ13964	NM_032186	G3A	NM_019101
AF140225	NM_030799	BAT1	NM_004640	CDW92	NM_080546	D13S106E	NM_005800	FAPP2	NM_032639	FLJ14153	NM_022736	G6PC	NM_000151
AF15Q14	NM_020380	BAT3	NM_004639	CEACAM1	NM_001712	D6S2654E	NM_012135	FBXL7	NM_012304	FLJ14393	NM_032778	G6PT1	NM_001467
AGA	NM_000027	BAT4	NM_033177	CEP3	NM_006449	DAF	NM_000574	FBXO24	NM_012172	FLJ14431	NM_032783	GAB1	NM_002039
AGM1	NM_015599	BAZ1A	NM_013448	CERD4	NM_012074	DAG1	NM_004393	FBXO4	NM_012176	FLJ14511	NM_033087	GABPA	NM_002040
AGPAT1	NM_006411	BAZ1B	NM_032408	CETN2	NM_004344	DBI	NM_020548	FBXO8	NM_012180	FLJ14821	NM_032811	GABPB2	NM_002041
AGT	NM_000029	BCAT2	NM_001190	CEZANNE	NM_020205	DBP	NM_001352	FBXW2	NM_012164	FLJ14824	NM_032819	GADD45G	NM_006705
AGXT2	NM_031900	BCCIP	NM_016567	CFL2	NM_021914	DBT	NM_001918	FDX1	NM_004109	FLJ14842	NM_032818	GAPD	NM_002046
AGXT2L1	NM_031279	BCDO2	NM_031938	CG005	NM_014887	DC11	NM_020186	FDXR	NM_024417	FLJ14981	NM_032824	GBE1	NM_000158
AHSG	NM_001622	BCL6	NM_001706	CGBP	NM_014593	DC13	NM_020188	FE65L2	NM_006051	FLJ14997	NM_032826	GC20	NM_005875
AK2	NM_001625	BCS1L	NM_004328	CGI-01	NM_015935	DC8	NM_015471	FEM1A	NM_020177	FLJ14827	NM_032846	GCHFR	NM_005258
AKAP13	NM_007200	BET1	NM_005868	CGI-11	NM_015941	DCK	NM_000788	FEM1B	NM_015322	FLJ14840	NM_032850	GCKR	NM_001486
AKR1C2	NM_001354	BF	NM_001710	CGI-51	NM_015380	DCLRE1B	NM_022836	FETUB	NM_014375	FLJ20010	NM_019021	GDAP2	NM_017686
AKR1C3	NM_003739	BHMT	NM_001713	CHD1L	NM_004284	DCLRE1C	NM_022487	FH	NM_000143	FLJ20014	NM_017622	GFER	NM_005262
AKR1C4	NM_001818	BIKE	NM_017593	CHI3L1	NM_001276	DDA3	NM_032636	FHIT	NM_002012	FLJ20037	NM_017633	GGCX	NM_000821
ALCAM	NM_001627	BIRC6	NM_016252	CHIC2	NM_012110	DDX18	NM_006773	FIGF	NM_004469	FLJ20080	NM_017657	GIOT-2	NM_016284
ALDH1A1	NM_000689	BLOV1	NM_018656	CHM	NM_000390	DDX27	NM_017895	FKSG87	NM_032029	FLJ20081	NM_017658	GIPC2	NM_017655
ALDH2	NM_000690	BPHL	NM_004332	CHP	NM_007236	DDX28	NM_018380	FLJ10038	NM_017976	FLJ20084	NM_017659	GJA4	NM_002060
ALDH3A1	NM_000691	BRCA1	NM_007295	CIAO1	NM_004804	DDX35	NM_021931	FLJ10111	NM_017999	FLJ20123	NM_017674	GJB1	NM_000166
ALDH3B1	NM_000694	BRD4	NM_014299	CISH	NM_013324	DDX38	NM_014003	FLJ10116	NM_018000	FLJ20125	NM_017676	GK001	NM_020198
ALDH5A1	NM_001080	BRIP1	NM_032043	CITED2	NM_006079	DDX8	NM_004941	FLJ10143	NM_018009	FLJ20130	NM_017681	GLYAT	NM_005838
ALDH8A1	NM_022568	BTD	NM_000060	CKAP1	NM_001281	DED	NM_012138	FLJ10276	NM_018045	FLJ20202	NM_017709	GMDS	NM_001500
ALDOC	NM_005165	BTF3	NM_001207	CKN1	NM_000082	DEDD2	NM_133328	FLJ10287	NM_019083	FLJ20287	NM_017746	GNB1L	NM_053004
ALS2	NM_020919	BTG1	NM_001731	CKS2	NM_001827	DEPP	NM_007021	FLJ10330	NM_018081	FLJ20331	NM_017768	GNG5	NM_005274
ALS2CR19	NM_057177	BTN2A1	NM_078476	CL683	NM_015696	DGKD	NM_003648	FLJ10407	NM_018087	FLJ20442	NM_017823	GNMT	NM_018960
AMACR	NM_014324	BYSL	NM_004053	CLCN3	NM_001829	DJ37E16.5	NM_020315	FLJ10415	NM_018089	FLJ20452	NM_017828	GNS	NM_002076
AMBP	NM_001633	C12orf8	NM_006817	CLCN6	NM_001286	DJ726C3.2	NM_025227	FLJ10422	NM_018091	FLJ20511	NM_017853	GOLGA2	NM_004488
AMOT	NM_133265	C14orf1	NM_007176	CLCNKA	NM_004070	DKFZP434C245	NM_015426	FLJ10432	NM_019070	FLJ20534	NM_017867	GOLGA4	NM_002078
AMT	NM_000481	C14orf3	NM_012111	CLDN2	NM_020384	DKFZp434D177	NM_032264	FLJ10482	NM_018107	FLJ20560	NM_017887	GOLPH4	NM_014498
ANG	NM_001145	C1orf8	NM_004872	CLDN3	NM_001306	DKFZP434H0115	NM_031421	FLJ10511	NM_018120	FLJ20595	NM_017894	GOSR2	NM_004287
ANKRA2	NM_023039	C1S	NM_001734	CLONE24922	NM_015679	DKFZP434J037	NM_030952	FLJ10525	NM_018126	FLJ20619	NM_017904	GOT1	NM_002079
ANPEP	NM_001150	C2	NM_000063	CLPTM1	NM_001294	DKFZP434L0117	NM_022778	FLJ10535	NM_018129	FLJ20627	NM_017905	GPC6	NM_005708
ANXA5	NM_001154	C20orf13	NM_017714	CLPX	NM_006660	DKFZP564A2416	NM_015535	FLJ10581	NM_018146	FLJ20628	NM_017910	GPHN	NM_020806
ANXA6	NM_001155	C20orf154	NM_032485	CLTA	NM_001833	DKFZP564G2022	NM_015497	FLJ10583	NM_018148	FLJ20671	NM_017924	GPR39	NM_001508
ANXA9	NM_003568	C20orf163	NM_080749	CLTCL1	NM_001835	DKFZP564L2423	NM_030805	FLJ10604	NM_018154	FLJ20699	NM_017931	GPT	NM_005309
AP1M1	NM_032493	C20orf164	NM_080752	CLYBL	NM_138280	DKFZP564O0463	NM_014156	FLJ10637	NM_018164	FLJ20707	NM_017936	GPX1	NM_000581
AP2A1	NM_130787	C20orf172	NM_024918	CNOT2	NM_014515	DKFZP564O0523	NM_032120	FLJ10640	NM_019023	FLJ20718	NM_017939	GPX2	NM_002083
AP3B1	NM_003664	C20orf188	NM_015638	CNOT4	NM_013316	DKFZP566C243	NM_015388	FLJ10661	NM_018172	FLJ20729	NM_017953	GRHPR	NM_012203
AP3M1	NM_012095	C20orf32	NM_020356	COASTER	NM_015555	DKFZP566M1046	NM_032127	FLJ10761	NM_018208	FLJ20730	NM_017945	GRIK3	NM_000831
AP4B1	NM_006594	C20orf4	NM_015511	COPB	NM_016451	DKFZp566O084	NM_015510	FLJ10774	NM_024662	FLJ21007	NM_030794	GRIN2D	NM_000836
APC10	NM_014885	C20orf84	NM_033550	COPB2	NM_004766	DKFZP586A011	NM_015416	FLJ10856	NM_018247	FLJ21144	NM_022774	GRO3	NM_002090
APCS	NM_001639	C20orf7	NM_024120	COPS7B	NM_022730	DKFZP586A0522	NM_014033	FLJ10871	NM_018250	FLJ21272	NM_025032	GSK3B	NM_002093
APEH	NM_001640	C20orf72	NM_052865	COQ3	NM_017421	DKFZP586J0119	NM_015636	FLJ10891	NM_018260	FLJ21415	NM_024738	GSPT1	NM_002094
APG-1	NM_014278	C20orf77	NM_021215	COX11	NM_004375	DKFZp762L0311	NM_018719	FLJ11000	NM_018295	FLJ21820	NM_021925	GSS	NM_000178
APG3	NM_022488	C21orf18	NM_017438	COX7A2	NM_001865	DLEU1	NM_005887	FLJ11011	NM_018299	FLJ21908	NM_024604	GSTA4	NM_001512
APMCF1	NM_021203	C21orf33	NM_004649	COX7A2L	NM_004718	DLST	NM_001933	FLJ11029	NM_018304	FLJ21934	NM_024743	GSTM4	NM_000850
APOA1	NM_000039	C2F	NM_005331	CPB2	NM_016413	DMC1	NM_007068	FLJ11046	NM_018309	FLJ21939	NM_022461	GSTTLp28	NM_004832
APOA2	NM_001643	C3	NM_000064	CPSF5	NM_007006	DNAJA2	NM_005880	FLJ11159	NM_018343	FLJ21963	NM_024560	GTF2E1	NM_005513

Fig. 18B

Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq
GTF2H1	NM_005316	IGF1	NM_000618	LOC51015	NM_016048	METAP2	NM_006838	MRPL49	NM_004927	OSGEP	NM_017807	PPP1R12B	NM_032105
GTPBG3	NM_032620	IGFBP1	NM_000596	LOC51016	NM_016049	MGAT4B	NM_054013	MRPL51	NM_016497	OSMR	NM_003999	PPP1R15B	NM_032833
GYS2	NM_021957	IL11RA	NM_004512	LOC51026	NM_016072	MGC13379	NM_016499	MRPS11	NM_022839	OTC	NM_000531	PPP1R3B	NM_024607
H2AFS	NM_080596	IL15	NM_000585	LOC51027	NM_016074	MGC10433	NM_024321	MRPS14	NM_022100	p100	NM_014390	PPP1R3C	NM_005398
H2AFG	NM_021065	IL1RAP	NM_002182	LOC51054	NM_015899	MGC10702	NM_032663	MRPS16	NM_016065	P115	NM_003715	PPP1R3D	NM_006242
H2AFO	NM_003516	IL22R	NM_021258	LOC51060	NM_015913	MGC10823	NM_031437	MRPS18B	NM_014046	p21UAS11	NM_000389	PPP2CA	NM_002715
H2BFA	NM_003518	IL2RB	NM_000878	LOC51064	NM_015917	MGC10924	NM_030571	MRPS18C	NM_016067	p21UAS8	NM_000389	PPP2R5B	NM_006244
H2BFB	NM_021063	IL6ST	NM_002184	LOC51074	NM_015957	MGC10940	NM_032303	MRPS21	NM_018997	p21UAS9	NM_000389	PPP4R1	NM_005134
H2BFF	NM_021062	IMMT	NM_006839	LOC51091	NM_016955	MGC10960	NM_032653	MRPS30	NM_016640	P23	NM_006601	PPP5C	NM_006247
H2BFG	NM_003522	INADL	NM_005799	LOC51096	NM_016001	MGC10974	NM_032306	MRPS35	NM_021821	P29	NM_015484	PQBP1	NM_005710
H326	NM_015726	INHBC	NM_005538	LOC51104	NM_016014	MGC10999	NM_032307	MRPS36	NM_033281	P2RY2	NM_002564	PRCC	NM_005973
H4F2	NM_003548	INVS	NM_014425	LOC51107	NM_016022	MGC11034	NM_031453	MRPS7	NM_015971	PABPC1	NM_002568	PRCP	NM_005040
H4FD	NM_003541	IRF6	NM_006147	LOC51134	NM_016122	MGC11266	NM_024322	MRS2L	NM_020662	PABPN1	NM_004643	PRKAB2	NM_005399
H6PD	NM_004285	ITGA6	NM_000210	LOC51142	NM_016139	MGC1127	NM_033549	MST1	NM_020998	PAFAH2	NM_000437	PRKCABP	NM_012407
HAAO	NM_012205	ITGAL	NM_002209	LOC51143	NM_016141	MGC11279	NM_024326	MSTP028	NM_031954	PAI-RBP1	NM_015640	PRKCL2	NM_006256
HADH2	NM_004493	ITIH3	NM_002217	LOC51174	NM_016261	MGC12435	NM_031427	MT1H	NM_005951	PAK4	NM_005884	PRLR	NM_000949
HADHA	NM_000182	ITIH4	NM_002218	LOC51175	NM_016262	MGC12943	NM_032317	MT1L	NM_002450	PALMD	NM_017734	PRO1728	NM_018505
HADHB	NM_000183	ITM1	NM_002219	LOC51187	NM_016304	MGC12981	NM_032357	MT1X	NM_005952	PANK	NM_138316	PRO2389	NM_025230
HADHSC	NM_005327	ITPR2	NM_002223	LOC51205	NM_016361	MGC13008	NM_032686	MT2A	NM_005953	PARVB	NM_013327	PRO2831	NM_018540
HAL	NM_002108	JIK	NM_016281	LOC51231	NM_016440	MGC13017	NM_080656	MTHFD1	NM_005956	PAX8	NM_013952	PROZ	NM_003891
HAO1	NM_017545	JRKL	NM_003772	LOC51240	NM_016457	MGC13033	NM_031447	MTHFR	NM_005957	PBEF	NM_005746	PRPF31	NM_015629
HARC	NM_017913	JUN	NM_002228	LOC51246	NM_016479	MGC13102	NM_032323	MTHFS	NM_006441	PCDH20	NM_022843	PRPS1	NM_002764
HAX1	NM_006118	JunB(-)1kb	NM_002229	LOC51285	NM_016563	MGC13138	NM_033410	MTMR2	NM_003912	PCK1	NM_002591	PRSS25	NM_013247
HBP1	NM_012257	JunB(-)2kb	NM_002229	LOC51287	NM_016565	MGC13159	NM_032927	MTMR4	NM_004687	PCK2	NM_004563	PSA	NM_021154
HBQ1	NM_005331	JunB(-)3kb	NM_002229	LOC51292	NM_016576	MGC1346	NM_032758	MTP	NM_000253	PCMT1	NM_005389	PSMA1	NM_002786
HBS1L	NM_006620	KAP3A	NG_000941	LOC51326	NM_016632	MGC14151	NM_032356	mtTFB	NM_016020	PCYT1A	NM_005017	PSMA2	NM_002787
HBXIP	NM_006402	KBRAS1	NM_020345	LOC51596	NM_015921	MGC14421	NM_032907	MUT	NM_000255	PDCD4	NM_014456	PSMA5	NM_002790
HCA112	NM_018487	KCNC3	NM_004977	LOC51601	NM_015929	MGC14433	NM_032904	MYO1A	NM_005379	PDE11A	NM_016953	PSMD10	NM_002814
HCDI	NM_020195	KCNJ12	NM_021012	LOC51611	NM_015958	MGC14839	NM_080659	N8AMT1	NM_013240	PDE4DIP	NM_014644	PSMD7	NM_002811
HDAC6	NM_006044	KCNN2	NM_021614	LOC51633	NM_016023	MGC14844	NM_032341	NAGA	NM_000262	PDE6D	NM_002601	PSME3	NM_005789
HDAC6	NM_006044	KEO4	NM_006459	LOC51644	NM_016057	MGC15435	NM_032367	NAGK	NM_017567	PDIR	NM_005810	PTD012	NM_014039
HEL308	NM_133636	KHDRBS1	NM_006559	LOC51651	NM_016077	MGC15504	NM_032751	NAPA	NM_003827	PDK2	NM_002611	PTD013	NM_015952
HEXA	NM_000520	KIAA0092	NM_014679	LOC51659	NM_016095	MGC15523	NM_138570	NAT8	NM_003960	PDK4	NM_002612	PTD015	NM_014040
HEY1	NM_012258	KIAA0102	NM_014752	LOC54516	NM_019041	MGC15563	NM_032876	NBP	NM_025233	PDZK1	NM_002614	PTK2	NM_005607
HFL3	NM_005666	KIAA0103	NM_014673	LOC54518	NM_019043	MGC15677	NM_032878	NCALD	NM_032041	PECI	NM_006117	PTPN18	NM_014369
HGC6.2	NM_014356	KIAA0105	NM_004906	LOC55580	NM_017571	MGC15737	NM_032926	NCBP1	NM_002486	PELO	NM_015946	PTPN4	NM_002830
HGD	NM_000187	KIAA0141	NM_014773	LOC55815	NM_018430	MGC15906	NM_032885	NCBP2	NM_007362	PEMT	NM_007169	PTPRE	NM_006504
HIF1A	NM_001530	KIAA0205	NM_014873	LOC55954	NM_019103	MGC16733	NM_033547	NCF1	NM_000265	PEPD	NM_000285	PTPRG	NM_002841
HINT2	NM_032593	KIAA0255	NM_014742	LOC56834	NM_020155	MGC16943	NM_080663	NCK1	NM_006153	PEX11B	NM_003846	PURG	NM_013357
HKE2	NM_014250	KIAA0268	NM_014785	LOC56902	NM_020143	MGC17347	NM_138333	NCOA3	NM_006534	PEX13	NM_002618	PWP1	NM_007062
HKE4	NM_006979	KIAA0266	NM_021645	LOC57018	NM_020307	MGC19595	NM_033415	NCOA5	NM_020967	PEX16	NM_057174	PYGL	NM_002863
HLA-B	NM_005514	KIAA0391	NM_014672	LOC57019	NM_020313	MGC2404	NM_032360	NCOR1	NM_006311	PEX3	NM_003630	PZP	NM_002864
HLA-F	NM_018950	KIAA0409	NM_015324	LOC57107	NM_020381	MGC2474	NM_023931	NDRG1	NM_006096	PFKFB4	NM_004567	QP-C	NM_014402
HMCS	NM_017947	KIAA0433	NM_015216	LOC57226	NM_020467	MGC2477	NM_024099	NDUFA4	NM_002489	PGM1	NM_002633	R3HDM	NM_015361
HMG1	NM_002128	KIAA0438	NM_014819	LOC57406	NM_020676	MGC2488	NM_024039	NDUFA6	NM_002490	PHACS	NM_032592	RA410	NM_016106
HMG17L3	NM_006353	KIAA0618	NM_014833	LOC57826	NM_021183	MGC2560	NM_031452	NDUFB1	NM_004545	PHLDA1	NM_007350	RAB10	NM_016131
HMOX2	NM_002134	KIAA0645	NM_014662	LOC57862	NM_021188	MGC2629	NM_032522	NDUFB5	NM_002492	PHTF1	NM_006608	RAB11A	NM_004663
HNF4a7	AF509467	KIAA0660	NM_012297	LOC64182	NM_022359	MGC2650	NM_024108	NDUFS1	NM_005006	PIGPC1	NM_022121	RAB18	NM_021252
HNMT	NM_006895	KIAA0670	NM_014977	LOC81034	NM_030780	MGC2734	NM_033117	NDUFS4	NM_002495	PIGPC1	NM_022121	RAB2	NM_002865
HNRPA1	NM_031157	KIAA0747	NM_015292	LOC81558	NM_030802	MGC2747	NM_024104	NEDD8	NM_006156	PIGPC1	NM_022121	RAB30	NM_014488
HNRPR	NM_005826	KIAA0792	NM_014698	LOC84518	NM_032488	MGC2835	NM_024072	NEK2	NM_002497	PIGPC1	NM_022121	RAB33B	NM_031296
HOXK3	NM_032410	KIAA0795	NM_025010	LOC84661	NM_032574	MGC3067	NM_024295	NET-2	NM_012338	PIGS	NM_033198	RAB4B	NM_016154
HOXA1	NM_005522	KIAA0806	NM_014813	LOC89953	NM_138343	MGC3180	NM_024041	NFE2L1	NM_003204	PIK3R3	NM_003629	RAB6KIFL	NM_005733
HOXC8	NM_022658	KIAA0872	NM_014940	LOC90799	NM_138363	MGC3222	NM_024334	NFKBIB	NM_002503	PIK4CB	NM_002651	RAB9P40	NM_005833
HPCL2	NM_012260	KIAA0905	NM_014933	LOC91689	NM_033318	MGC3248	NM_032486	NFKBIB	NM_002503	PILB	NM_012228	RABEX5	NM_014504
HPN	NM_002151	KIAA0914	NM_014883	LR8	NM_014020	MGC3413	NM_032678	NFKBIB	NM_002503	PINK1	NM_032409	RAD17	NM_133338
HPRP4P	NM_004697	KIAA1017	NM_007216	LSM3	NM_014463	MGC4161	NM_024303	NFKBIB	NM_002503	PIP5K1A	NM_003557	RAD23B	NM_002874
HPX	NM_000613	KIAA1041	NM_014947	LSR7	NM_018559	MGC4189	NM_032306	NFYA	NM_002505	PIPOX	NM_016518	RAD50	NM_133482
HRIHFB2436	NM_014345	KIAA1116	NM_014892	LTA4H	NM_000895	MGC4400	NM_032679	NKTR	NM_005385	PIR	NM_003662	RAGA	NM_006570
HSA011916	NM_015343	KIAA1169	NM_017901	LZTR1	NM_006767	MGC4606	NM_024516	NME1	NM_000269	PIST	NM_020399	RA-GEF-2	NM_016340
HSD11B1	NM_005525	KIAA1453	NM_025090	M17S2	NM_031858	MGC4638	NM_031479	NOLC1	NM_004741	PITPNB	NM_012399	RAMP	NM_016448
HSD17B2	NM_002153	KIAA1638	NM_025132	M96	NM_007358	MGC4683	NM_024514	NONO	NM_007363	PKM2	NM_002654	RANBP8	NM_006390
HSD17B4	NM_000414	KIF1B	NM_015074	MADCAM1	NM_007164	MGC4677	NM_052871	NPAS2	NM_002518	PLA2G13	NM_032562	RANGAP1	NM_002883
HSD17B7	NM_016371	KIF9	NM_022342	MADH4	NM_005359	MGC4767	NM_032314	NPAT	NM_002519	PLAB	NM_004864	RAP1GA1	NM_002885
HSPA5	NM_005347	KLF15	NM_014079	MAF	NM_005360	MGC5302	NM_024089	NPC1	NM_000271	PLAGL2	NM_002657	RASA1	NM_022650
HSPC002	NM_015362	KLHL6	NM_130446	MAGOH	NM_002370	MGC5509	NM_024093	NR0B2	NM_021969	PLD2	NM_002663	RASSF1	NM_007182
HSPC048	NM_014148	KNG	NM_000893	MAL2	NM_052886	MGC9084	NM_033418	NR1H3	NM_005693	PLGL	NM_002665	RBBP4	NM_005610
HSPC051	NM_013387	KNSL4	NM_007317	MANBA	NM_005908	MGEA5	NM_012215	NR1I2	NM_022002	PLSCR1	NM_021105	RBM15	NM_022768
HSPC052	NM_014150	KPNB1	NM_002265	MAOA	NM_000240	MGST1	NM_020300	NR3C1	NM_000176	PME-1	NM_016147	RBM6	NM_005777
HSPC111	NM_016391	KRT10	NM_000421	MAP3K11	NM_002419	MGST2	NM_002413	NR5A2	NM_003822	PMS1	NM_000534	RBM7	NM_016090
HSPC117	NM_014306	LAD1	NM_005558	MAP3K4	NM_005922	MGST3	NM_004528	NRAS	NM_002524	PMS2	NM_000535	RBP5	NM_031491
HSPC129	NM_016396	LALP1	NM_020354	MAP3K7	NM_003188	MIPEP	NM_005932	NRCAM	NM_005010	PMS2L8	NM_005394	RBSK	NM_022128
HSPC141	NM_014172	LAPTM4A	NM_014713	MAP7	NM_003980	MLC1SA	NM_002475	NRD1	NM_002525	PNAS-131	NM_031446	RBT1	NM_013368
HSPC154	NM_014177	LATS1	NM_004690	MAPK7	NM_002749	MNAT1	NM_002431	NS1-BP	NM_006469	PNKP	NM_007254	RCL	NM_006443
HSPC157	NM_014179	LBP	NM_004139	MAT1A	NM_000429	MOV10	NM_020963	NT5C3	NM_016489	PNLIPRP1	NM_006229	RDBP	NM_002904
HSPC165	NM_014186	LC27	NM_018407	MAT2A	NM_005911	MPP1	NM_002436	NTHL1	NM_002528	POLB	NM_002690	RDH5	NM_002905
HSPC213	NM_016475	LCN2	NM_005564	MBD4	NM_003925	MPPE1	NM_023075	NTN4	NM_021229	POLD4	NM_021173	REA	NM_007273
HSU79274	NM_013390	LENG5	NM_024075	MCCC1	NM_020166	MRE11A	NM_005590	NUDT2	NM_001161	POLE3	NM_017443	RECQL5	NM_004259
HSU84971	NM_013303	LEPR	NM_002303	MCEE	NM_032601	MRP63	NM_024026	NUDT5	NM_014142	POLR2A	NM_000937	RENT1	NM_002911
HT002	NM_014066	LGALS1	NM_002305	MCP	NM_002389	MRPL15	NM_014175	NUFIP1	NM_012345	POLR2K	NM_005034	RFC3	NM_002915
HT007	NM_018480	LIMK2	NM_005569	MDFI	NM_005586	MRPL18	NM_014161	NUP107	NM_020401	POLS	NM_006999	RFC5	NM_007370
HT010	NM_018471	LISCH7	NM_015925	MDH1	NM_005917	MRPL2	NM_015950	NUP62	NM_012346	PON1	NM_000446	RGL	NM_015149
HT012	NM_018473	LIV-1	NM_012319	MDM2UAS6	NM_002392	MRPL24	NM_024540	NUP98	NM_005387	POP5	NM_015918	RIG-I	NM_014314
humNRDR	NM_021004	LNPEP	NM_005575	MDM2UAS8	NM_002392	MRPL33	NM_004891	OAS1	NM_002534	PORIMIN	NM_052932	RIP60	NM_013400
HYAL3	NM_003549	LOC115330	NM_138445	MDS009	NM_020234	MRPL34	NM_023937	OAS3	NM_006187	POV1	NM_003627	RNASE2	NM_002934
IER5	NM_016545	LOC129401	NM_138285	MDS025	NM_021825	MRPL37	NM_016491	OAZ2	NM_002537	PPS395	NM_021732	RNASE3	NM_002935
IFITM2	NM_006435	LOC151534	NM_138482	MDS029	NM_018464	MRPL4	NM_015956	OPA3	NM_025136	PPFIBP1	NM_003622	RNASE4	NM_002937
IFNAR1	NM_000629	LOC151636	NM_138287	MEA	NM_014823	MRPL44	NM_022915	ORC3L	NM_012381	PPGB	NM_000308	RNF29	NM_033058
IFNGR1	NM_000416	LOC51004	NM_015940	MEF2B	NM_005919	MRPL46	NM_022162	ORM1	NM_000607	PPM1D	NM_003620	RNF5	NM_006913
IFRD1	NM_001550	LOC51011	NM_016044	MEP50	NM_024102	MRPL48	NM_016055	ORM2	NM_000608	PPP1R11	NM_021959	RNGTT	NM_003800

Fig. 18C

Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq
RNPC2	NM_004902	SLC25A13	NM_014251	TDRKH	NM_006862	VPS45A	NM_007259
RNPEPL1	NM_018226	SLC25A18	NM_031481	TEAD3	NM_003214	VTN	NM_000638
ROCK1	NM_005406	SLC25A5	NM_001152	TED	NM_015686	WASF3	NM_006646
RORC	NM_005060	SLC26A1	NM_022042	TEF	NM_003216	WASL	NM_003941
RPC32	NM_006467	SLC2A8	NM_014580	TEGT	NM_003217	WBP4	NM_007187
RPL18	NM_000979	SLC31A1	NM_001859	TESK2	NM_007170	WDF2	NM_052950
RPL31	NM_000993	SLC35A2	NM_005660	TF	NM_001063	WDR10	NM_052985
RPL37AP1	NG_000988	SLC35A3	NM_012243	THPO	NM_000460	WDR12	NM_018256
RPL5	NM_000969	SLC38A1	NM_030674	THTP	NM_024328	WDR13	NM_017883
RPL7	NM_000971	SLC38A4	NM_018018	TIA1	NM_022037	XDH	NM_000379
RPLP1	NM_001003	SLC39A1	NM_014437	TIMM17A	NM_006335	XPA	NM_000380
RPS16	NM_001020	SLC5A3	NM_006933	TIMM17B	NM_005834	XPC	NM_004628
RPS19	NM_001022	SLC7A2	NM_003046	TIMM23	NM_006327	XPR1	NM_004736
RPS27A	NM_002954	SLC7A9	NM_014270	TIMM9	NM_012460	XRCC5	NM_021141
RPS3A	NM_001006	SLPI	NM_003064	TLH29	NM_032036	YKT6	NM_006555
RPS6KA5	NM_004755	SMAC	NM_019887	TLN1	NM_006289	YWHAB	NM_003404
RPS6KB1	NM_003161	SMAP	NM_006696	TM4SF4	NM_004617	ZAN	NM_003386
RQCD1	NM_005444	SMARCA5	NM_003601	TM9SF2	NM_004800	ZBRK1	NM_021632
RSHL1	NM_030785	SMARCE1	NM_003079	TMEM7	NM_031440	ZF5128	NM_014347
RSP3	NM_031924	SMC2L1	NM_006444	TMF1	NM_007114	ZFP95	NM_014569
RSU1	NM_012425	SMPD1	NM_000543	TMOD2	NM_014548	ZK1	NM_005815
RTCD1	NM_003729	SNAI2	NM_003068	TMP21	NM_006827	ZNF133	NM_003434
RTP801	NM_019058	SNAP23	NM_003825	TNFAIP1	NM_021137	ZNF144	NM_007144
RUVBL2	NM_006666	SNAPC1	NM_003082	TNFRSF11B	NM_002546	ZNF146	NM_007145
RXRB	NM_021976	SNK	NM_006622	TNFRSF6	NM_000043	ZNF147	NM_005082
S100A9	NM_002965	SNRPA	NM_004596	TNFRSF6	NM_000043	ZNF155	NM_003445
SAA1	NM_000331	SNRPD3	NM_004175	TNFRSF6	NM_000043	ZNF183	NM_006978
SAA1	NM_000331	SNRPF	NM_003095	TNFRSF6	NM_000043	ZNF192	NM_006298
SAA1	NM_000331	SNW1	NM_012245	TNFSF13	NM_003808	ZNF207	NM_003457
SAA1	NM_000331	SNX1	NM_003099	TNS	NM_022648	ZNF214	NM_013249
SAA2	NM_030754	SNX17	NM_014748	TOM1	NM_005488	ZNF22	NM_006963
SAC	NM_018417	SNX3	NM_003795	TOMM70A	NM_014820	ZNF221	NM_013359
SAD1	NM_006590	SNX5	NM_014426	TP53TG1	NM_007233	ZNF222	NM_013360
SAMHD1	NM_015474	SOD1	NM_000454	TPP2	NM_003291	ZNF224	NM_013398
SAP18	NM_005870	SORCS3	NM_014978	TPT	NM_014317	ZNF225	NM_013362
SAS10	NM_020368	SOX10	NM_006941	TRA1	NM_003299	ZNF226	NM_016444
SC4MOL	NM_006745	SP2	NM_138406	TRAF6	NM_004620	ZNF230	NM_006300
SCA2	NM_002973	SPATA2	NM_006038	TRAP150	NM_005119	ZNF237	NM_014242
SCAND1	NM_033630	SPATA6	NM_019073	TRIM15	NM_033229	ZNF281	NM_012482
SCD	NM_005063	SPC18	NM_014300	TRIM26	NM_003449	ZNF302	NM_018443
SCYA14	NM_032962	SPOCK	NM_004598	TRIM31	NM_052816	ZNF361	NM_018555
SCYA15	NM_032964	SPP2	NM_006944	TRIM34	NM_130389	ZNF9	NM_003418
SCYA16	NM_004590	SQRDL	NM_021199	TRIM4	NM_033017	ZNF-U69274	NM_014415
SCYE1	NM_004757	SREBF2	NM_004599	TRIP11	NM_004239	ZNRD1	NM_014596
SDC1	NM_002997	SRP54	NM_003136	TRN-SR	NM_012470	ZnTL2	NM_133496
SDCCAG10	NM_005869	SRP68	NM_014230	TRPC5	NM_012471		
SDCCAG28	NM_006645	SRPR	NM_003139	TRPS1	NM_014112		
SDFR1	NM_012428	SSA2	NM_004600	TSG101	NM_006292		
SEC10L1	NM_006544	SSAT2	NM_133491	TSLRP	NM_012472		
SEC23A	NM_006364	SSSCA1	NM_006396	TTY14	NM_031932		
SEC24D	NM_014822	SSTR1	NM_001049	TUBB5	NM_006087		
SEC61B	NM_006808	STAF42	NM_053053	TUFT1	NM_020127		
SEL1L	NM_005065	STAF65(gamma)	NM_014860	TXNIP	NM_006472		
SEMA3C	NM_006379	STAM	NM_003473	TXNL	NM_004786		
SEMA6C	NM_030913	STAM2	NM_005843	TXNRD1	NM_003330		
SEMA7A	NM_003612	STARD7	NM_020151	TYMS	NM_001071		
SENP1	NM_014554	STAT1	NM_007315	U2AF1	NM_006758		
SEPX1	NM_016332	STAU2	NM_014393	U3-55K	NM_004704		
SERPINA1	NM_000295	STCH	NM_006948	U5-116KD	NM_004247		
SERPINA10	NM_016186	STIM1	NM_003156	UBE2B	NM_003337		
SERPINA5	NM_000624	STK19	NM_004197	UBE2D3	NM_003340		
SERPINA6	NM_001756	STK2	NM_003157	UBE2M	NM_003969		
SERPINC1	NM_000488	STOML1	NM_004809	UBP1	NM_014517		
SERPIND1	NM_000185	STRAIT11499	NM_021242	UBQLN1	NM_053067		
SERPINE1	NM_000602	STX18	NM_016930	UBQLN2	NM_013444		
SERPING1	NM_000062	SUCLA2	NM_003850	UCH37	NM_015984		
SERPINI1	NM_005025	SUCLG1	NM_003849	UCHL3	NM_006002		
SES2	NM_031459	SUDD	NM_003831	UGDH	NM_003359		
SF3A3	NM_006802	SULT1A1	NM_001055	UGT2B11	NM_001073		
SF3B2	NM_006842	SULT2A1	NM_003167	UGT2B15	NM_001076		
SFRS11	NM_004768	SUOX	NM_000456	UGTREL1	NM_005827		
SFRS5	NM_006925	SUPT3H	NM_003599	UGTREL7	NM_015139		
SFRS8	NM_004592	SUPT5H	NM_003169	ULBP3	NM_024518		
SGK	NM_005627	SUPV3L1	NM_003171	UPB1	NM_016327		
SGK2	NM_016276	SYN3	NM_133632	UQCRC2	NM_003366		
SGT1	NM_006704	SYTL4	NM_080737	URKL1	NM_017859		
SH2D3C	NM_005489	SZF1	NM_016089	UROD	NM_000374		
SH3BGRL2	NM_031469	TADA3L	NM_133480	UROS	NM_000375		
SILV	NM_006928	TAF2GL	NG_001012	USP1	NM_003368		
SIX2	NM_016932	TAGLN2	NM_003564	USP15	NM_006313		
SKB1	NM_006109	TARS	NM_003191	USP2	NM_004205		
SKD1	NM_004869	TAT	NM_000353	UXT	NM_004182		
SKRP1	NM_080876	TCF1	NM_000545	VAMP1	NM_014231		
SLC10A1	NM_003049	TCF12	NM_003205	VAMP5	NM_006634		
SLC17A2	NM_005835	TCF19	NM_007109	VDAC1	NM_003374		
SLC17A5	NM_012434	TCF21	NM_003206	VDAC2	NM_003375		
SLC19A3	NM_025243	TCF7L2	NM_030756	VEGFC	NM_005429		
SLC22A1LS	NM_007105	TCIRG1	NM_006019	VEZATIN	NM_017599		
SLC22A3	NM_021977	TCOF1	NM_000356	VMP1	NM_030938		
SLC22A7	NM_006672	TCP1	NM_030752	VPS29	NM_016226		

Fig. 19A

Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq
101F6	NM_007022	BIG1	NM_006421	CGBP	NM_014593	DKFZP547N043	NM_032018	FLJ10477	NM_018105	FLJ20420	NM_017812	GPRK2L	NM_005307
4E-T	NM_019843	BLTR2	NM_018839	CGI-01	NM_015935	DKFZP564G2022	NM_015497	FLJ10509	NM_018119	FLJ20422	NM_017814	GRIK3	NM_000831
AAMP	NM_001087	BLZF1	NM_003666	CGI-203	NM_020408	DKFZP564I0422	NM_031435	FLJ10511	NM_018120	FLJ20450	NM_017827	GRTH	NM_013284
ABCB10	NM_012089	BM-002	NM_016617	CGI-51	NM_015380	DKFZP564L2423	NM_030805	FLJ10525	NM_018126	FLJ20498	NM_019040	GRWD	NM_031485
ABCB8	NM_007188	BMI1	NM_005180	CHEK2	NM_007194	DKFZP564M082	NM_014042	FLJ10535	NM_018129	FLJ20508	NM_017850	GSPT1	NM_002094
ABCB9	NM_019624	BMP5	NM_021073	CHERP	NM_006387	DKFZP564O0463	NM_014156	FLJ10581	NM_018145	FLJ20511	NM_017853	GSS	NM_000178
ABCC5	NM_005688	BNC	NM_001717	CHIC2	NM_012110	DKFZP564O0523	NM_032120	FLJ10583	NM_018148	FLJ20546	NM_017872	GSTZ1	NM_001513
ABCG1	NM_004915	BNIP1	NM_001205	CHM	NM_000390	DKFZP566B183	NM_015509	FLJ10604	NM_018154	FLJ20558	NM_017880	GTF2B	NM_001514
ABH	NM_006020	BPGM	NM_001724	CHMP1.5	NM_020412	DKFZP566C243	NM_015388	FLJ10628	NM_018159	FLJ20624	NM_017906	GTF2E1	NM_005513
ABS	NM_016222	BRAP	NM_006768	CHRNB2	NM_000748	DKFZP566D1346	NM_030816	FLJ10634	NM_018163	FLJ20627	NM_017909	GTF2H1	NM_005316
ABT1	NM_013375	BRCA1	NM_007295	CIAO1	NM_004804	DKFZP566E144	NM_015523	FLJ10637	NM_018164	FLJ20628	NM_017910	GTF2H3	NM_001516
ACAD8	NM_014384	BRF2	NM_018310	CIP29	NM_032364	DKFZP586A011	NM_015416	FLJ10640	NM_019023	FLJ20643	NM_017916	GTF2H4	NM_001517
ACADSB	NM_001609	BRIX	NM_018321	CIR	NM_004882	DKFZP586J0119	NM_015636	FLJ10661	NM_018172	FLJ20644	NM_017917	GTF2I	NM_033003
ACATN	NM_004733	BST1	NM_004334	CITED2	NM_006079	DKFZP761E2110	NM_030953	FLJ10774	NM_024682	FLJ20651	NM_017919	GTF3C5	NM_012087
ACO2	NM_001098	BTD	NM_000060	CKAP1	NM_001281	DKFZp781J139	NM_032280	FLJ10803	NM_018224	FLJ20671	NM_017924	GUSB	NM_000181
ACOX1	NM_004035	BTRC	NM_033637	CKS2	NM_001827	DKFZP762H166	NM_020441	FLJ10826	NM_018233	FLJ20695	NM_017929	H. GS165L1	NM_004904
ACOX3	NM_003501	BUB1B	NM_001211	CLLD8	NM_031915	DLG4	NM_001365	FLJ10853	NM_018246	FLJ20729	NM_017953	H17	NM_017547
ACP2	NM_001610	BUB3	NM_004725	CLONE24922	NM_015679	DMAP1	NM_019100	FLJ10856	NM_018247	FLJ20730	NM_017945	H326	NM_015726
ACTR1A	NM_005736	BYSL	NM_004053	CLPTM1	NM_001294	DMP1	NM_004407	FLJ10871	NM_018250	FLJ20731	NM_017946	H3FM	NM_021059
AD-017	NM_018446	C11orf10	NM_014206	CLPX	NM_006660	DNAJB11	NM_016306	FLJ10891	NM_018260	FLJ20748	NM_019020	H4F2	NM_003548
AD022	NM_016514	C11orf2	NM_013265	CLTA	NM_001833	DNAJB12	NM_017626	FLJ10989	NM_018292	FLJ20772	NM_017956	H4FI	NM_003544
AD034	NM_031480	C14orf3	NM_012111	CLTCL1	NM_001835	DNAJB4	NM_007034	FLJ10998	NM_018294	FLJ20859	NM_022734	HAAO	NM_012205
AD158	NM_032270	C1D	NM_006333	CNAP1	NM_014865	DPAGT1	NM_001382	FLJ11000	NM_018295	FLJ21272	NM_025032	HACAP-G	NM_022346
AD24	NM_022451	C1orf22	NM_025191	CNOT3	NM_014516	DYRK1B	NM_004714	FLJ11193	NM_018356	FLJ21952	NM_022494	HCDI	NM_020195
ADAT1	NM_012091	C1orf25	NM_030934	CNOT4	NM_013316	E2F4	NM_001950	FLJ11220	NM_018304	FLJ21977	NM_032213	HCNGP	NM_013260
ADCY7	NM_001114	C1orf8	NM_004872	COASTER	NM_015555	E2F5	NM_001951	FLJ11271	NM_018373	FLJ21986	NM_024913	HD	NM_002111
ADD2	NM_001617	C20orf1	NM_012112	COPG	NM_006710	E2IG3	NM_014366	FLJ11274	NM_018375	FLJ22028	NM_024854	HDAC8	NM_018486
ADSS	NM_001126	C20orf10	NM_014477	COPB	NM_016451	EAF1	NM_033083	FLJ11292	NM_018382	FLJ22169	NM_024085	HEC	NM_006101
AF093680	NM_013242	C20orf111	NM_016470	COPB2	NM_004766	EED	NM_003797	FLJ11301	NM_018385	FLJ22184	NM_025094	HEL308	NM_133636
AF140225	NM_030789	C20orf12	NM_018152	COPS7A	NM_016319	EEF1B2	NM_021121	FLJ11838	NM_024654	FLJ22191	NM_025231	HEXA	NM_000520
AF15Q14	NM_020380	C20orf13	NM_017714	COPS7B	NM_022730	EFG1	NM_024996	FLJ11848	NM_025155	FLJ22347	NM_022830	HGD	NM_000187
AGA	NM_000027	C20orf14	NM_012469	COX7A2	NM_001865	E2F5	NM_001951	FLJ11271	NM_018373	FLJ22501	NM_024747	HHEX	NM_022729
AGTPBP1	NM_015239	C20orf154	NM_032485	COX7A2L	NM_004718	E2IG3	NM_014366	FLJ11274	NM_018375	FLJ22551	NM_024708	HHLA2	NM_007072
AIP	NM_003977	C20orf164	NM_080752	COX7C	NM_001867	EAF1	NM_033083	FLJ11292	NM_018382	FLJ22637	NM_025165	HIF1AN	NM_017902
AK2	NM_001625	C20orf188	NM_015638	COX8	NM_004074	EED	NM_003797	FLJ11301	NM_018385	FLJ22688	NM_025129	HIRIP3	NM_003609
AKR1B1	NM_001628	C20orf28	NM_015417	CPA2	NM_001869	EEF1B2	NM_021121	FLJ11838	NM_024654	FLJ22729	NM_024683	HKE2	NM_014260
ALS2	NM_020919	C20orf30	NM_014145	CPSF5	NM_007006	EFG1	NM_024996	FLJ11848	NM_025155	FLJ22865	NM_025109	HKE4	NM_006979
AMSH	NM_006463	C20orf33	NM_030877	CPT1B	NM_004377	EGLN2	NM_053046	FLJ12085	NM_022771	FLJ23109	NM_024814	HLF	NM_002126
AMXRA2	NM_023039	C20orf4	NM_015511	CREBL1	NM_004381	EHD3	NM_014600	FLJ12188	NM_024682	FLJ23182	NM_022366	HMG1	NM_002128
AP1M1	NM_032493	C20orf43	NM_016407	CREBL2	NM_001310	EIF1A	NM_001412	FLJ12455	NM_022078	FLJ23251	NM_024818	HMG2	NM_002129
AP2A1	NM_130787	C20orf44	NM_018244	CRFG	NM_012341	EIF2B1	NM_001414	FLJ12525	NM_031206	FLJ23263	NM_025115	HNRPA0	NM_006805
AP2B1	NM_001282	C20orf45	NM_016045	CrkRS	NM_016507	EIF2S1	NM_004094	FLJ12577	NM_024926	FLJ23468	NM_024629	HNRPA1	NM_031157
AP2M1	NM_004068	C20orf64	NM_033550	CRSP3	NM_004830	EIF2S2	NM_003908	FLJ12707	NM_022067	FLJ23469	NM_024710	HNRPC	NM_031314
AP2S1	NM_021575	C20orf72	NM_052866	CRY2	NM_001889	EIF2S3	NM_001415	FLJ12735	NM_024857	FLJ23499	NM_022761	HPCL2	NM_012260
AP3M1	NM_012095	C20orf77	NM_021215	CRYZL1	NM_005111	EIF3S2	NM_003757	FLJ12770	NM_032174	FLJ23875	NM_032231	HPRP4P	NM_004697
AP4B1	NM_006594	C21orf18	NM_017438	CS	NM_004077	EIF3S6	NM_001568	FLJ12785	NM_024855	FLJ23879	NM_024814	HRB2	NM_007043
APACD	NM_005783	C21orf55	NM_017833	CSK	NM_004383	EIF4G1	NM_004953	FLJ12788	NM_022492	FLJ23879	NM_022366	HRMT1L2	NM_001536
APC10	NM_014885	C21orf59	NM_021254	CSNK2A1	NM_001895	EIF5	NM_001969	FLJ12879	NM_024757	FLJ23879	NM_024818	HSGT1	NM_007265
APG3	NM_022488	C2F	NM_006331	CSTF1	NM_001324	ELL	NM_005532	FLJ12888	NM_024945	FLJ23879	NM_025115	HSP105B	NM_006644
APMCF1	NM_021203	C2orf9	NM_032309	CSTF2T	NM_015235	EPHA1	NM_005232	FLJ12895	NM_023926	FLJ23879	NM_024629	HSPA5	NM_005347
AQP3	NM_004925	C3orf4	NM_019895	CSTF3	NM_001326	ERCC5	NM_000123	FLJ12910	NM_024573	FLJ23879	NM_024710	HSPCO03	NM_014017
AQP6	NM_001652	C4orf1	NM_005345	CTAG1	NM_001327	EWSR1	NM_013986	FLJ12960	NM_024638	FLJ23879	NM_022761	HSPCO16	NM_015933
ARD1	NM_003491	C5orf6	NM_016605	CTMP	NM_053055	EXO1	NM_130398	FLJ13102	NM_024887	FLNA	NM_001456	HSPCO31	NM_016101
ARF1GAP	NM_018209	C6orf11	NM_005452	CTNNA1	NM_001903	EZFIT	NM_021216	FLJ13158	NM_024909	FNTB	NM_002028	HSPCO51	NM_013387
ARFD1	NM_001656	C6orf35	NM_018452	CUL2	NM_003591	F12	NM_000505	FLJ13194	NM_025146	FOXO1A	NM_002015	HSPCO52	NM_014150
ARHGAP11.	NM_014783	C7orf10	NM_024728	CXorf12	NM_003492	F23149_1	NM_019088	FLJ13195	NM_022906	FOXO1A	NM_002015	HSPCO56	NM_014154
ARL1	NM_001177	C9orf12	NM_022755	CYB5-M	NM_030579	FACTP140	NM_007192	FLJ13220	NM_021927	FRG1	NM_004477	HSPCO72	NM_014162
ARS2	NM_015908	C9orf5	NM_032012	CYLD	NM_015247	FANCF	NM_022725	FLJ13273	NM_024751	FRSB	NM_005887	HSPCO16	NM_015933
ARSDR1	NM_016026	CAP	NM_006367	CYP51	NM_000786	FBXO24	NM_012172	FLJ13291	NM_032178	FTL	NM_000146	HSPCO31	NM_016101
ASB3	NM_016116	CAPZA2	NM_006136	D123	NM_006023	FBXO6	NM_012180	FLJ13315	NM_025005	FTSJ1	NM_012280	HSPCO51	NM_013387
ASE-1	NM_012099	CAT56	NM_025263	D13S106E	NM_005800	FBXW2	NM_012164	FLJ13491	NM_024623	FUBP1	NM_003902	HSPCO52	NM_014150
ATF4	NM_001675	CAV1	NM_001753	D1S155E	NM_007158	FDPS	NM_002004	FLJ13611	NM_024941	FXC1	NM_021192	HSPCO56	NM_014154
ATF6	NM_007348	CBARA1	NM_006077	DACH2	NM_053281	FDX1	NM_004109	FLJ13615	NM_025114	FYCO1	NM_024513	HSPCO72	NM_014162
ATF7	NM_006856	CBX5	NM_012117	DAD1	NM_001344	FDXR	NM_024417	FLJ13798	NM_024773	G10	NM_003910	HSPC111	NM_016391
ATP10C	NM_024490	CCNE1	NM_001238	DBT	NM_001918	FE65L2	NM_006051	FLJ13912	NM_022770	G22P1	NM_001469	HSPC117	NM_014306
ATP5B	NM_001686	CCNT1	NM_001240	DC11	NM_020186	FEN1	NM_004111	FLJ13949	NM_025077	G6PD	NM_000402	HSPC128	NM_014157
ATP5F1	NM_001688	CCT6B	NM_006584	DC13	NM_020188	FGF13	NM_004114	FLJ13962	NM_024682	GABPA	NM_002040	HSPC129	NM_016396
ATP5G3	NM_001689	CCT7	NM_006429	DC50	NM_031210	FGF7	NM_002009	FLJ14431	NM_032783	GABPB2	NM_002041	HSPC134	NM_014169
ATP5J2	NM_004889	CCT8	NM_006585	DC8	NM_015471	FH	NM_000143	FLJ14451	NM_032786	GABRE	NM_021984	HSPC138	NM_016401
ATP6E	NM_001696	CDC10	NM_001788	DCLRE1B	NM_022836	FHIT	NM_002012	FLJ14486	NM_032792	GALNAC4S	NM_031422	HSPC141	NM_014172
ATP6M	NM_015994	CDC23	NM_004661	DCTN4	NM_016221	FKBP10	NM_021939	FLJ14511	NM_033087	GAS1	NM_002048	HSPC142	NM_014173
ATP6S14	NM_004231	CDC25A	NM_001789	DDOST	NM_005216	FKBP3	NM_002013	FLJ14547	NM_032804	GBF1	NM_004193	HSPC144	NM_014174
AUP1	NM_012103	CDC42BPB	NM_006035	DDX10	NM_004398	FKBPL	NM_022110	FLJ14697	NM_032826	GCN5L1	NM_001487	HSPC148	NM_016403
AUTL1	NM_032852	CDC45L	NM_003504	DDX21	NM_004728	FKSG32	NM_031307	FLJ14803	NM_032842	GDAP2	NM_017686	HSPC152	NM_016404
B3GNT6	NM_006876	CDC5L	NM_001253	DDX28	NM_018380	FLJ10038	NM_017679	FLJ14840	NM_032850	GHITM	NM_014394	HSPC157	NM_014179
BAD	NM_004322	CDC6	NM_001254	DDX38	NM_014003	FLJ10052	NM_017682	FLJ14855	NM_033210	GIOT-3	NM_016265	HSPC160	NM_014182
BAG4	NM_004874	CDCA1	NM_031423	DDX8	NM_004941	FLJ10116	NM_018000	FLJ20010	NM_019021	GJA4	NM_002060	HSPC166	NM_014186
BARD1	NM_000465	CDIPT	NM_006319	DED	NM_012138	FLJ10142	NM_018008	FLJ20045	NM_017638	GK001	NM_020198	HSPC171	NM_014187
BAT1	NM_004640	CDK5	NM_004935	DEDD	NM_004216	FLJ10276	NM_018045	FLJ20070	NM_017652	GLA	NM_000169	HSPC182	NM_014188
BAT2	NM_004638	CDK8	NM_001260	DESC1	NM_014058	FLJ10287	NM_019083	FLJ20080	NM_017657	GLTSCR2	NM_015710	HSPC189	NM_016535
BAT3	NM_004639	CDKN1B	NM_004064	DGUOK	NM_080915	FLJ10330	NM_018061	FLJ20081	NM_017658	GNAI3	NM_006496	HSPE1	NM_002157
BAT4	NM_033177	CEBPA	NM_004364	DIS3	NM_014953	FLJ10342	NM_018064	FLJ20084	NM_017659	GNB2L1	NM_006098	HSU79274	NM_013300
BAZ1B	NM_032408	CEBPB	NM_005194	DJ37E16.5	NM_020315	FLJ10374	NM_018074	FLJ20125	NM_017676	GNS	NM_002076	HSU84971	NM_013303
BCAR1	NM_014567	CEP2	NM_006779	DKFZP434B1(	NM_015434	FLJ10377	NM_018077	FLJ20189	NM_017704	GOSR1	NM_004871	HT010	NM_018471
BCCIP	NM_016567	CES2	NM_003869	DKFZP434C2(	NM_015426	FLJ10407	NM_018087	FLJ20190	NM_017705	GOSR2	NM_004287	HT011	NM_018472
BCKDHA	NM_000709	CETN2	NM_004344	DKFZp434E22	NM_017612	FLJ10415	NM_018089	FLJ20257	NM_019606	GOT1	NM_002079	HUNK	NM_014586
BCL2L1	NM_001191	CETN3	NM_004365	DKFZP434E2:	NM_032138	FLJ10422	NM_018091	FLJ20268	NM_024668	GPCR150	NM_014373	IER5	NM_016545
BCS1L	NM_004328	CFL1	NM_005507	DKFZP434L11	NM_032146	FLJ10432	NM_019070	FLJ20291	NM_017748	GPR105	NM_014879	IFRD1	NM_001550
BET1	NM_005868	CG005	NM_014887	DKFZp434N0E	NM_032261	FLJ10450	NM_018095	FLJ20342	NM_017774	GPR37	NM_005302	IFRD2	NM_006764
BET3	NM_014408	CG11	NM_006349	DKFZp434N14	NM_032133	FLJ10468	NM_018101	FLJ20343	NM_017775	GPR52	NM_005884	IGBP1	NM_001551

Fig. 19B

Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq	Gene Name	RefSeq
IGSF6	NM_052868	LOC51075	NM_015959	MFAP1	NM_005926	MRPL33	NM_004891	NTE	NM_006702	PPP1R15B	NM_032833	RPL7	NM_000971
IMAGE145052	NM_014267	LOC51076	NM_015960	MGC10433	NM_024321	MRPL43	NM_032112	NTF2	NM_005796	PPP2R5B	NM_006244	RPLP0L	NM_016183
IMAGE345520	NM_024006	LOC51077	NM_015962	MGC10500	NM_031477	MRPL44	NM_022915	NUCB1	NM_006184	PPP6C	NM_002721	Rpo1-2	NM_019014
IMMT	NM_006839	LOC51094	NM_015999	MGC10702	NM_032663	MRPL46	NM_022163	NUDT2	NM_001161	PRCC	NM_005973	RPS14	NM_005617
IMP13	NM_014652	LOC51096	NM_016001	MGC10924	NM_030571	MRPL48	NM_016055	NUDT5	NM_014142	PRDM5	NM_018699	RPS16	NM_001020
INCENP	NM_020238	LOC51104	NM_016014	MGC10974	NM_032306	MRPL51	NM_016497	NUDT6	NM_007083	PRDX5	NM_012094	RPS18	NM_022551
ING3	NM_019071	LOC51107	NM_016022	MGC10999	NM_032307	MRPL53	NM_053050	NUP107	NM_020401	PRKAB1	NM_006253	RPS19	NM_001022
ING4	NM_016162	LOC51117	NM_016035	MGC11102	NM_032225	MRPS11	NM_022839	NUP54	NM_017426	PRKCABP	NM_012407	RPS20	NM_001023
INVS	NM_014425	LOC51118	NM_016037	MGC11115	NM_032310	MRPS12	NM_021107	NUP62	NM_012346	PRKCE	NM_005400	RPS21	NM_001024
IRS4	NM_003604	LOC51142	NM_016139	MGC11266	NM_024322	MRPS14	NM_022100	NVL	NM_002533	PRO2389	NM_025230	RPS25	NM_001028
ITGA6	NM_000210	LOC51174	NM_016261	MGC1127	NM_033549	MRPS15	NM_031280	NYD-SP11	NM_031951	PRP18	NM_003675	RPS27A	NM_002954
ITGA9	NM_002207	LOC51187	NM_016304	MGC11279	NM_024326	MRPS16	NM_016065	NY-REN-41	NM_080654	PRPF31	NM_015629	RPS28	NM_001031
ITGB3BP	NM_014288	LOC51202	NM_016355	MGC11296	NM_032352	MRPS18B	NM_014046	OBTP	NM_013397	PRRG2	NM_000951	RPS3	NM_001005
ITM1	NM_002219	LOC51204	NM_016360	MGC11352	NM_030927	MRPS18C	NM_016067	OGFR	NM_007346	PRSS25	NM_013247	RPS3A	NM_001006
JM4	NM_007213	LOC51205	NM_016361	MGC12943	NM_032317	MRPS21	NM_018997	OPA1	NM_015560	PSCD2	NM_004226	RPS5	NM_001009
JTB	NM_006694	LOC51231	NM_016440	MGC12981	NM_032357	MRPS23	NM_016070	OPA3	NM_025136	PSMA1	NM_002786	RPS6	NM_001010
KARS	NM_005548	LOC51246	NM_016479	MGC13102	NM_032323	MRPS27	NM_015084	ORC1L	NM_004153	PSMA2	NM_002787	RPS6KA5	NM_004755
KBRAS1	NM_020345	LOC51287	NM_016565	MGC13114	NM_032366	MRPS28	NM_014018	ORC3L	NM_012381	PSMA3	NM_002788	RPS6KB1	NM_003161
KCNQ5	NM_019842	LOC51290	NM_016570	MGC13138	NM_033410	MRPS30	NM_016640	OSBP	NM_002556	PSMA5	NM_002790	RPS8KC1	NM_012424
KEO4	NM_006459	LOC51292	NM_016576	MGC13159	NM_032927	MRPS35	NM_021821	OSBPL11	NM_022776	PSMB1	NM_002793	RRM1	NM_001033
KIAA0026	NM_015340	LOC51300	NM_016589	MGC1346	NM_032758	MRPS7	NM_015971	OSCAR	NM_130771	PSMB5	NM_002797	RRP4	NM_014285
KIAA0057	NM_012288	LOC51326	NM_016632	MGC14126	NM_032898	MSMB	NM_002443	OSGEP	NM_017807	PSMB7	NM_002799	RRP46	NM_020158
KIAA0092	NM_014879	LOC51329	NM_016636	MGC14151	NM_032356	MSTP028	NM_031954	P125	NM_007190	PSMC4	NM_006503	RSU1	NM_012425
KIAA0102	NM_014752	LOC51596	NM_015921	MGC14288	NM_032901	MTERF	NM_006980	P15-2	NM_018698	PSMD1	NM_002807	RXRB	NM_021976
KIAA0105	NM_004906	LOC51604	NM_015937	MGC14421	NM_032907	MTF1	NM_005955	P29	NM_015484	PSMD10	NM_002814	SACM2L	NM_022553
KIAA0184	NM_014739	LOC51605	NM_015939	MGC14595	NM_032334	MTHFD1	NM_005956	P5326	NM_031450	PSMD4	NM_002810	SAD1	NM_005590
KIAA0196	NM_014846	LOC51626	NM_016008	MGC14697	NM_032747	MTMR4	NM_004687	PACE	NM_002569	PSMD7	NM_002811	SAP18	NM_005870
KIAA0255	NM_014742	LOC51631	NM_016019	MGC14798	NM_080650	MTR	NM_000254	PACE4	NM_002570	PSMD8	NM_002812	SART3	NM_014706
KIAA0258	NM_014785	LOC51633	NM_016023	MGC14835	NM_033412	MTRF1	NM_004294	PAFAH2	NM_000437	PSME3	NM_005789	SAS10	NM_020368
KIAA0274	NM_014845	LOC51644	NM_016057	MGC15677	NM_032878	mtTFB	NM_016020	PAI-RBP1	NM_015640	PTD009	NM_016146	SBP2	NM_024077
KIAA0317	NM_014821	LOC51651	NM_016077	MGC16169	NM_033115	MUT	NM_000255	PANX2	NM_052839	PTD012	NM_014039	SCDGF-B	NM_025206
KIAA0372	NM_014639	LOC51657	NM_016086	MGC16386	NM_080668	MUTYH	NM_012222	PAPA-1	NM_031288	PTD013	NM_015952	SCML1	NM_006746
KIAA0391	NM_014672	LOC51691	NM_016200	MGC16733	NM_033547	MXI1	NM_005962	PARVB	NM_013327	PTD015	NM_014040	SCYE1	NM_004757
KIAA0416	NM_015564	LOC54516	NM_019041	MGC17347	NM_138333	MYCBP	NM_012333	PAWR	NM_002583	PTK7	NM_002821	SDCCAG10	NM_005869
KIAA0419	NM_014711	LOC54543	NM_019059	MGC19595	NM_033415	MYL6	NM_079424	PAX1	NM_006192	PTPN13	NM_006264	SDCCAG28	NM_006645
KIAA0426	NM_014724	LOC55815	NM_018430	MGC20486	NM_052844	NAG	NM_015909	pcno	NM_020357	PWP1	NM_007062	SDF2	NM_006923
KIAA0433	NM_015216	LOC55954	NM_019103	MGC2404	NM_032360	NAGK	NM_017567	PCQAP	NM_015889	R3HDM	NM_015361	SDFR1	NM_012428
KIAA0438	NM_014819	LOC56851	NM_020154	MGC2408	NM_032331	NAKAP95	NM_014371	PCYT1A	NM_005017	RA410	NM_016106	SDHC	NM_003001
KIAA0537	NM_014840	LOC56902	NM_020143	MGC24447	NM_138288	NAPA	NM_003827	PDCD10	NM_007217	RAB11A	NM_004663	SEC10L1	NM_006544
KIAA0547	NM_014793	LOC56993	NM_020243	MGC2474	NM_023931	NBP	NM_025233	PDE4DIP	NM_014644	RAB18	NM_021252	SEC22L1	NM_004892
KIAA0670	NM_014977	LOC57019	NM_020313	MGC2477	NM_024099	NBR2	NM_005821	PDE6D	NM_002601	RAB1B	NM_030981	SEC3	NM_018261
KIAA0682	NM_014852	LOC57107	NM_020381	MGC2488	NM_024039	NCBP1	NM_002486	PDE9A	NM_002606	RAB2	NM_002865	SEC61B	NM_006808
KIAA0710	NM_014871	LOC57109	NM_020385	MGC2508	NM_024327	NCBP2	NM_007362	PEAS	NM_057161	RAB30	NM_014488	SEC8	NM_021807
KIAA0766	NM_014805	LOC57147	NM_020423	MGC2650	NM_031452	NCOA4	NM_005437	PEF	NM_012392	RAB6KIFL	NM_005733	SEDLP	NM_015890
KIAA0795	NM_025010	LOC63929	NM_022098	MGC2650	NM_024108	NDUFA1	NM_004541	PEMT	NM_007169	RAB7	NM_004637	SEL1L	NM_005065
KIAA0806	NM_014813	LOC81034	NM_030780	MGC2655	NM_024339	NDUFA3	NM_004542	PET112L	NM_004564	RABAC1	NM_006423	SENP1	NM_014554
KIAA0872	NM_014940	LOC81558	NM_030802	MGC2747	NM_024104	NDUFA4	NM_002489	PEX11B	NM_003846	Rabin4R	NM_017987	SERPINA4	NM_006215
KIAA0907	NM_014949	LOC88953	NM_138343	MGC2840	NM_024079	NDUFA5	NM_005000	PEX12	NM_000286	RAD51	NM_133487	SERPINB3	NM_006919
KIAA0950	NM_012306	LOC90346	NM_138351	MGC3121	NM_024031	NDUFA6	NM_002490	PEX13	NM_002618	RAGA	NM_006570	SERPINB8	NM_002640
KIAA0971	NM_014929	LOC90678	NM_138361	MGC3123	NM_024107	NDUFA7	NM_005001	PEX16	NM_057174	RAI2	NM_021785	SERPINI1	NM_005025
KIAA1012	NM_014939	LOC90701	NM_033280	MGC3133	NM_031287	NDUFB3	NM_002491	PEX6	NM_000287	RAMP	NM_016448	SES2	NM_031459
KIAA1017	NM_007216	LOC90799	NM_138363	MGC3180	NM_024041	NDUFB5	NM_002492	PFDN5	NM_002624	RANBP8	NM_006390	SETDB1	NM_012432
KIAA1041	NM_014947	LOC92106	NM_138381	MGC3222	NM_024334	NDUFS1	NM_005006	PHACS	NM_032592	RANGAP1	NM_002883	SF3A3	NM_006802
KIAA1068	NM_015332	LRP5	NM_002335	MGC3248	NM_032486	NDUFS3	NM_004551	PHB	NM_002634	RAP1	NM_018975	SF3B1	NM_012433
KIAA1100	NM_014901	LRRN1	NM_002319	MGC4054	NM_024341	NDUFS4	NM_002495	PHKB	NM_000293	RARG-1	NM_016167	SF3B2	NM_006842
KIAA1608	NM_024820	LSM3	NM_014463	MGC4093	NM_030578	NDUFV1	NM_007103	PIGN	NM_012327	RASSF1	NM_007182	SF3B4	NM_005850
KIAA1775	NM_033100	LSM4	NM_012321	MGC4161	NM_024303	NEDD8	NM_006156	PIGPC1	NM_022121	RBAK	NM_021163	SFRS1	NM_006924
KIF3B	NM_004798	LSM5	NM_012322	MGC4189	NM_032308	NEK7	NM_133494	PIGPC1	NM_022121	RBBP4	NM_005610	SFRS11	NM_004768
KIF9	NM_022342	LTA4H	NM_000895	MGC4251	NM_032376	NFATC2	NM_012340	PIGPC1	NM_022121	RBL1	NM_002895	SFRS2	NM_003016
KLRF1	NM_016523	LYPLA2	NM_007260	MGC4308	NM_032359	NFE2L1	NM_003204	PIK3C3	NM_002647	RBL2	NM_005611	SFRS5	NM_006925
KNSL7	NM_020242	LZTFL1	NM_020347	MGC4608	NM_024516	NFE2L3	NM_004289	PINK1	NM_032409	RBM15	NM_022768	SFRS6	NM_004592
KPTN	NM_007059	LZTR1	NM_006767	MGC4767	NM_032314	NFKBIB	NM_002503	PIP5K1A	NM_003557	RBM6	NM_005777	SGCE	NM_003919
KRT10	NM_000421	M17S2	NM_031858	MGC4771	NM_032668	NFKBIB	NM_002503	PIST	NM_020399	RBM7	NM_016090	SGT1	NM_006704
LAPTM4A	NM_014713	M6A	NM_019852	MGC5302	NM_024089	NFKBIB	NM_002503	PL6	NM_002994	RDBP	NM_002904	SH3BGRL2	NM_031469
LCMT	NM_016015	M9	NM_013234	MGC5347	NM_024063	NFKBIB	NM_002503	PLA2GZD	NM_012400	RDH5	NM_002905	SHH	NM_000193
LCP	NM_014315	M96	NM_007356	MGC5378	NM_032632	NFKBIL1	NM_005007	PLA2G4B	NM_005090	REA	NM_007273	SIP	NM_014412
LDB1	NM_003893	MAGOH	NM_002370	MGC5469	NM_032361	NFYA	NM_002505	PLAA	NM_004253	RECQL	NM_002907	SIRT2	NM_030593
LEPR	NM_002303	MAP2K5	NM_002757	MGC5509	NM_024093	NIMP	NM_032730	PLDN	NM_012388	RECQL5	NM_004259	SKB1	NM_006109
LGMN	NM_005606	MAP3K11	NM_002419	MGC5521	NM_024061	NKTR	NM_005385	PME-1	NM_016147	REG1A	NM_002909	SKD1	NM_004869
LHX6	NM_014368	MAP3K3	NM_002401	MGC9084	NM_033418	NLN	NM_020726	PMS2	NM_000535	REG1B	NM_006507	SKD3	NM_030813
LIM	NM_006457	MAP3K7	NM_003188	MGC9740	NM_080658	NMA	NM_012342	PMS2L8	NM_005394	RENT1	NM_002911	SKI	NM_003036
LIMS1	NM_004987	MAPK7	NM_002749	MGST3	NM_004528	NME1	NM_000269	PNAS-131	NM_031446	RFC3	NM_002915	SKP2	NM_032637
LIN-7-C	NM_018362	MAPK8IP2	NM_012324	MID1	NM_000381	NME7	NM_013330	PNKP	NM_007254	RFPL2	NM_005605	SLC16A6	NM_004694
LISCH7	NM_015925	MAPK8IP3	NM_015133	MKRN1	NM_013446	NOH61	NM_019082	PNMA1	NM_006029	RNF40	NM_014771	SLC25A19	NM_021734
LIV-1	NM_012319	MAT2A	NM_005911	MLH1	NM_000249	NOLA1	NM_018983	PODXL	NM_005397	RNF5	NM_006913	SLC2A8	NM_014580
LOC113251	NM_052879	MBD4	NM_003925	MLN	NM_002418	NOLC1	NM_004741	POLE3	NM_017443	RNGTT	NM_003800	SLC31A1	NM_001859
LOC113444	NM_138428	MCEE	NM_032601	MN1	NM_002430	NOSIP	NM_015953	POLL	NM_013274	RNPC2	NM_004902	SLC35A1	NM_006416
LOC113622	NM_138430	MCFP	NM_016843	MOCS3	NM_014484	NOT56L	NM_005787	POLR2A	NM_000937	RPA2	NM_002946	SLC35A2	NM_005660
LOC115827	NM_138453	MCM3	NM_002388	MPPE1	NM_023075	NPAS2	NM_002518	POLR2K	NM_005034	RPA40	NM_004575	SLC7A9	NM_014270
LOC129401	NM_138285	MDFI	NM_005586	MRE11A	NM_005590	NPC1	NM_000271	POLR3F	NM_006466	RPL10	NM_032241	SMAC	NM_019887
LOC151534	NM_138482	MDH1	NM_005917	MRPL11	NM_016050	NPR2L	NM_006545	POLRMT	NM_005035	RPL12	NM_000976	SMAP	NM_006898
LOC153768	NM_138492	MDH2	NM_005918	MRPL18	NM_014161	NR1D1	NM_021724	POP5	NM_015918	RPL18	NM_000979	SMARCA5	NM_003601
LOC51002	NM_016058	MDS025	NM_021825	MRPL19	NM_014763	NR1H3	NM_005693	POR1	NM_012402	RPL18A	NM_000980	SMARCE1	NM_003079
LOC51004	NM_015940	MDS032	NM_018457	MRPL2	NM_015950	NRAS	NM_002524	POU5F1	NM_002701	RPL26	NM_000987	SMC1L1	NM_006306
LOC51016	NM_016049	MDS033	NM_018468	MRPL22	NM_014180	NRCAM	NM_005010	PPIL1	NM_016059	RPL27	NM_000988	SMC2L1	NM_006444
LOC51019	NM_016053	MEF2B	NM_005919	MRPL24	NM_024540	NRD1	NM_002525	PPIL2	NM_014337	RPL31	NM_000993	SMC4L1	NM_005496
LOC51026	NM_016072	MEN1	NM_130800	MRPL27	NM_016504	NS1-BP	NM_006469	PPP1CA	NM_002708	RPL32	NM_000994	SMCX	NM_004187
LOC51027	NM_016074	MEP50	NM_024102	MRPL3	NM_007208	NSEP1	NM_004559	PPP1R10	NM_002714	RPL37	NM_000997	SMPD2	NM_003080
LOC51050	NM_015913	METAP2	NM_006838	MRPL30	NM_016503	NSF	NM_006178	PPP1R11	NM_021959	RPL37A	NM_000998	SNRPA	NM_004596
LOC51057	NM_015936	METL	NM_018396	MRPL32	NM_031903	NT5C3	NM_016489	PPP1R12B	NM_032105	RPL41	NM_021104	SNRPD2	NM_004597

Fig. 19C

Gene Name	RefSeq	Gene Name	RefSeq
SNRPD3	NM_004175	TXNL	NM_004
SNRPF	NM_003095	U2AF1	NM_006
SNW1	NM_012245	U5-100K	NM_004
SNX1	NM_003099	U5-116KD	NM_004
SNX11	NM_013323	UBE2M	NM_003
SNX17	NM_014748	UBE2N	NM_003
SNX5	NM_014426	UBE2V1	NM_022
SON	NM_003103	UBQLN1	NM_053
SOX17	NM_022454	UCH37	NM_016
SOX9	NM_000346	UGTREL1	NM_005
SP2	NM_138406	UMPS	NM_000
SPATA2	NM_006038	UNRIP	NM_007
SPC18	NM_014300	UPF3B	NM_080
SPG4	NM_014946	UQCRC2	NM_003
SPK	NM_004819	UQCRH	NM_006
SQRDL	NM_021199	URKL1	NM_017
SRP19	NM_003135	UROD	NM_000
SRP54	NM_003136	UROS	NM_000
SRP68	NM_014230	USF1	NM_007
SSA2	NM_004600	USP5	NM_003
SSBP1	NM_003143	UXT	NM_004
SSFA2	NM_006751	V1RL1	NM_020
SSR2	NM_003145	VEGFC	NM_005
SSR3	NM_007107	VMP1	NM_030
SSSCA1	NM_006396	VPS33A	NM_022
SSTK	NM_032037	WARS2	NM_015
SSTR4	NM_001052	WBP4	NM_007
ST13	NM_003932	WDF2	NM_052
STAF42	NM_053053	WDR12	NM_018
STAF65(gamı	NM_014860	WDR13	NM_017
STAM	NM_003473	WHIP	NM_020
STAM2	NM_005843	XPC	NM_004
STCH	NM_006948	XPO1	NM_003
STK19	NM_004197	XRCC4	NM_022
STK24	NM_003576	XRCC5	NM_021
STOML1	NM_004809	XRN2	NM_012
STOML2	NM_013442	YR-29	NM_014
STX18	NM_016930	YWHAB	NM_003
SUCLG1	NM_003849	ZBRK1	NM_021
SULT1A3	NM_003166	ZF5128	NM_014
SULT1C1	NM_001056	ZFP37	NM_003
SUPT5H	NM_003169	ZFP93	NM_004
SUPV3L1	NM_003171	ZFP95	NM_014
T54	NM_015698	ZNF133	NM_003
TADA3L	NM_133480	ZNF134	NM_003
TAF11	NM_005643	ZNF142	NM_005
TAF6	NM_005641	ZNF146	NM_007
TARBP2	NM_004178	ZNF155	NM_003
TAX1BP1	NM_006024	ZNF175	NM_007
TCERG1	NM_006706	ZNF183	NM_006
TCF1	NM_000545	ZNF189	NM_003
TCF2	NM_000458	ZNF192	NM_006
TCF2	NM_000458	ZNF193	NM_006
TCF2	NM_000458	ZNF207	NM_003
TCOF1	NM_000356	ZNF214	NM_013
TCP1	NM_030752	ZNF221	NM_013
TDRKH	NM_006862	ZNF222	NM_013
TEGT	NM_003217	ZNF224	NM_013
TESK2	NM_007170	ZNF225	NM_013
TFAP4	NM_003223	ZNF226	NM_016
TFPT	NM_013342	ZNF230	NM_006
TG737	NM_006531	ZNF264	NM_003
TIMM23	NM_006327	ZNF265	NM_005
TIMM9	NM_012460	ZNF277	NM_021
TIP39	NM_012143	ZNF300	NM_052
TLE3	NM_005078	ZNF302	NM_018
TLN1	NM_006289	ZNF304	NM_020
TM9SF1	NM_006405	ZNF317	NM_020
TM9SF2	NM_004800	ZNF338	NM_022
TMOD2	NM_014548	ZNF345	NM_003
TMP21	NM_006827	ZNF361	NM_018
TMSB10	NM_021103	ZNF-U69274	NM_014
TNFAIP1	NM_021137	ZNRD1	NM_014
TOMM70A	NM_014820		
TOR2A	NM_130459		
TPT	NM_014317		
TRA1	NM_003299		
TRAF5	NM_004619		
TRAP150	NM_005119		
TRFP	NM_004275		
TRIM4	NM_033017		
TRIP	NM_005879		
TRIP11	NM_004239		
TRN-SR	NM_012470		
TRPS1	NM_014112		
TSG101	NM_006292		
TSLRP	NM_012472		
TSN	NM_004622		
TSNAX	NM_005999		
TUBB4	NM_006086		

Fig. 20A

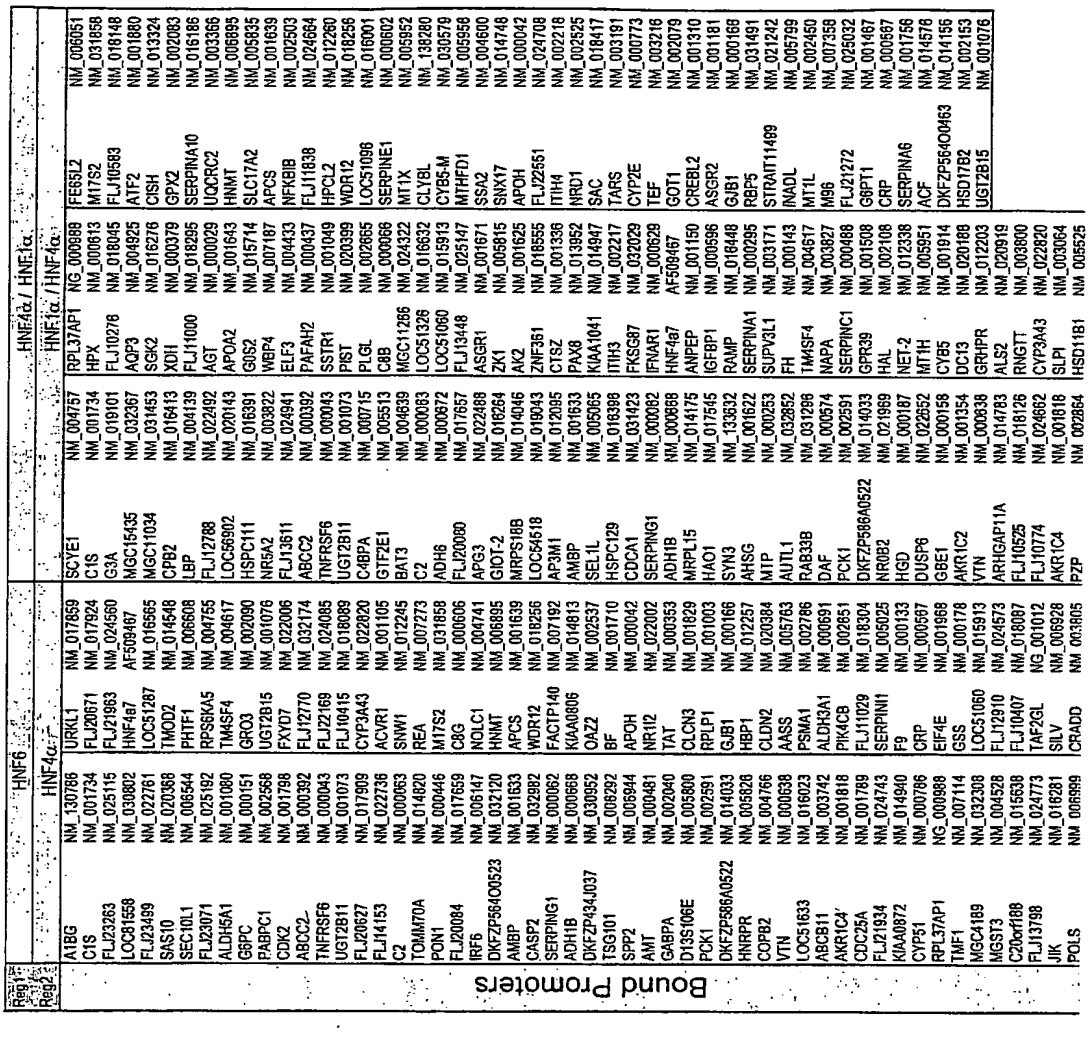


Fig. 20B

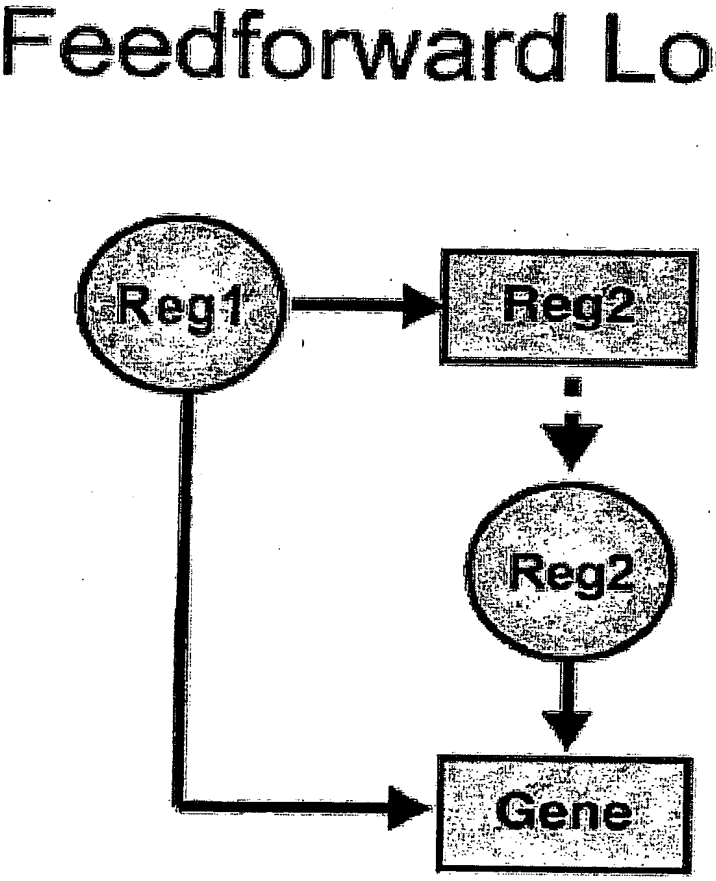


Fig. 21A

Reg1 Reg2 Reg3	HNF6 HNF4α HNF1α		HNF6 HNF1α	
Bound Promoters	C1S	NM_001734	F11	NM_019559
	ABCC2	NM_000392	C1S	NM_001734
	TNFRSF6	NM_000043	FLJ10650	NM_018168
	UGT2B11	NM_001073	ABCC2	NM_000392
	C2	NM_000063	TNFRSF6	NM_000043
	AMBP	NM_001633	UGT2B11	NM_001073
	SERPING1	NM_000062	UGT1A1	NM_000463
	ADH1B	NM_000668	C2	NM_000063
	PCK1	NM_002591	ADH1A	NM_000667
	DKFZP586A0522	NM_014033	AMBP	NM_001633
	VTN	NM_000638	SERPING1	NM_000062
	AKR1C4	NM_001818	ADH1B	NM_000668
	FLJ21934	NM_024743	HABP2	NM_004132
	KIAA0872	NM_014940	PCK1	NM_002591
	RPL37AP1	NG_000988	DKFZP586A	NM_014033
	PLGL	NM_002665	VTN	NM_000638
	C8B	NM_000066	AKR1C4	NM_001818
	LOC51060	NM_015913	FLJ21934	NM_024743
	HNF4a7	AF509467	KIAA0872	NM_014940
	TM4SF4	NM_004617	RPL37AP1	NG_000988
	UGT2B15	NM_001076	PLGL	NM_002665
	CYP3A43	NM_022820	C8B	NM_000066
	M17S2	NM_031858	LOC51060	NM_015913
	HNMT	NM_006895	HNF4a7	AF509467
	APCS	NM_001639	TM4SF4	NM_004617
	WDR12	NM_018256	UGT2B15	NM_001076
	APOH	NM_000042	CYP3A43	NM_022820
	GJB1	NM_000166	M17S2	NM_031858
	CRP	NM_000567	HNMT	NM_006895
			APCS	NM_001639
			WDR12	NM_018256
			APOH	NM_000042
			GJB1	NM_000166
			CRP	NM_000567

Fig. 21B

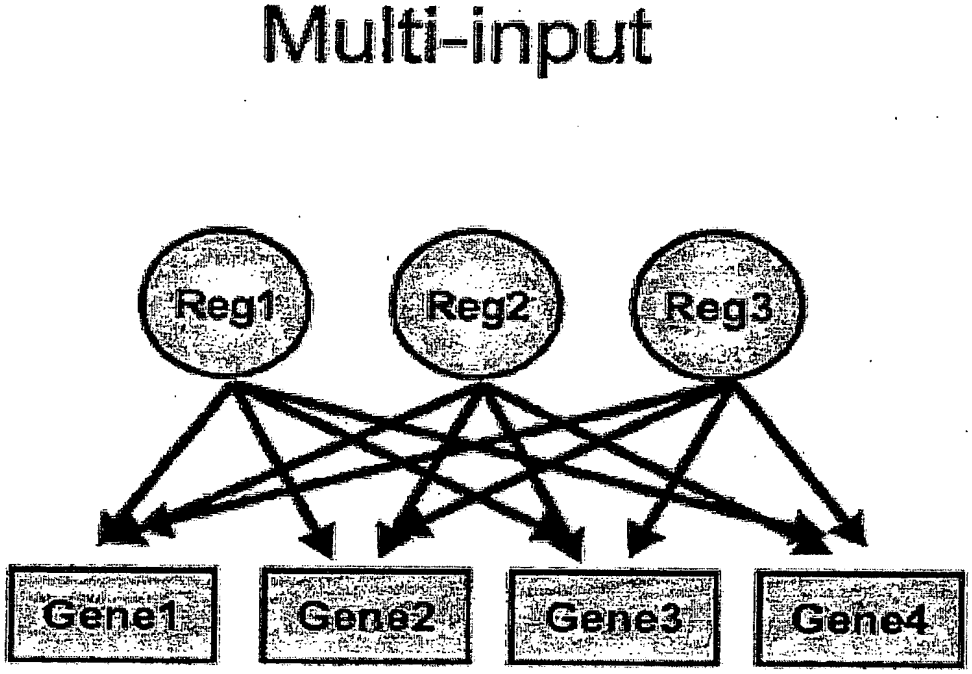


Fig. 22A

Reg1 / Reg2	HNF6 / HNF4α				HNF1α / HNF4α / HNF4α / HNF1α	
Bound Promoters	BCKDHA	NM_000709	FLJ13798	NM_024773	FLJ13273	NM_024751
	FLJ23263	NM_025115	GSS	NM_000178	MGC10500	NM_031477
	FLJ11271	NM_018373	HBOA	NM_007067	SDCCAG10	NM_005869
	HMG2	NM_002129	LOC51060	NM_015913	FBXO8	NM_012180
	LOC81558	NM_030802	FLJ13220	NM_021927	ZNF300	NM_052860
	SAS10	NM_020368	FLJ12910	NM_024573	H4F2	NM_003548
	SEC10L1	NM_006544	FLJ10407	NM_018087	FLJ11301	NM_018385
	RRP46	NM_020158	FLJ10342	NM_018084	SEL1L	NM_005065
	SNRPD2	NM_004597	FLJ20671	NM_017924	ZNF155	NM_003445
	MDH1	NM_005917	LOC51287	NM_016565	C6orf11	NM_005452
	ORC1L	NM_004153	GLA	NM_000169	ARHGAP11A	NM_014783
	FLJ20627	NM_017909	RPS6KA5	NM_004755	UROD	NM_000374
	GTF2E1	NM_005513	FLJ20772	NM_017956	FLJ20731	NM_017946
	TOMM70A	NM_014820	FLJ12770	NM_032174	RAB6KIFL	NM_005733
	PAPA-1	NM_031288	FLJ22169	NM_024085	TMP21	NM_006827
	HASJ4442	NM_017528	FLJ10415	NM_018089	MGC15677	NM_032878
	FLJ20084	NM_017659	ZNF317	NM_020933	WBP4	NM_007187
	PEX6	NM_000287	SNW1	NM_012245	PAFAH2	NM_000437
	FLJ11301	NM_018385	REA	NM_007273	EIF3S6	NM_001568
	EED	NM_003797	C2F	NM_006331	PSMA5	NM_002790
	MGC18595	NM_033415	NOLC1	NM_004741	TMOD2	NM_014548
	CIR	NM_004882	CLONE24922	NM_015679	GLA	NM_000169
	CLLD8	NM_031915	CCT8	NM_006585	GNB2L1	NM_006098
	ABCB8	NM_007188	PSMB1	NM_002793	FNTB	NM_002028
	SPG4	NM_014946	WDR12	NM_018256	PEX13	NM_002618
	GABPA	NM_002040	KIAA0806	NM_014813	FE65L2	NM_006051
	OGFR	NM_007346	DKFZp761J139	NM_032280	UQCRC2	NM_003366
	COPB2	NM_004766	SART3	NM_014706	FLJ14855	NM_033210
	AF15Q14	NM_020380	COX7A2L	NM_004718	HHLA2	NM_007072
	MTERF	NM_006980	FLJ20422	NM_017814	CYB5-M	NM_030579
	LOC51633	NM_016023	COPS7A	NM_016319	CDC45L	NM_003504
	FLJ14486	NM_032792	FLJ20643	NM_017916	pcnp	NM_020357
	FLJ21934	NM_024743	HBP1	NM_012257	FLJ20643	NM_017916
	KIAA0872	NM_014940	PSMA1	NM_002786	FLJ21272	NM_025032
	TEGT	NM_003217	FLJ21272	NM_025032		
	MGC4189	NM_032308	FLJ11029	NM_018304		
	SERPINB8	NM_002640	ARL1	NM_001177		
	MGST3	NM_004528	SERPINI1	NM_005025		
	HSP105B	NM_006644	NUDT2	NM_001161		
	C20orf188	NM_015638				

Table S11. The feed forward regulatory motifs in pancreatic islets . The regulatory modules here were derived as described in Supporting Online Material. Feed forwards only involving HNF1α and HNF4α are also multi-input motifs, as they bind each other's promoters in a multicomponent loop.

Fig. 22B

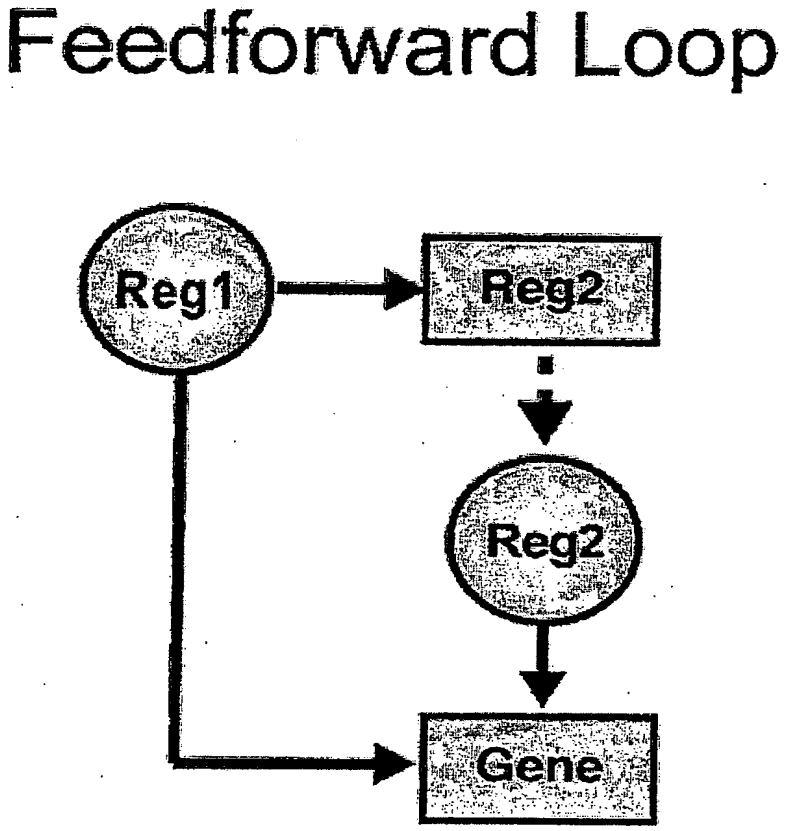


Fig. 23A

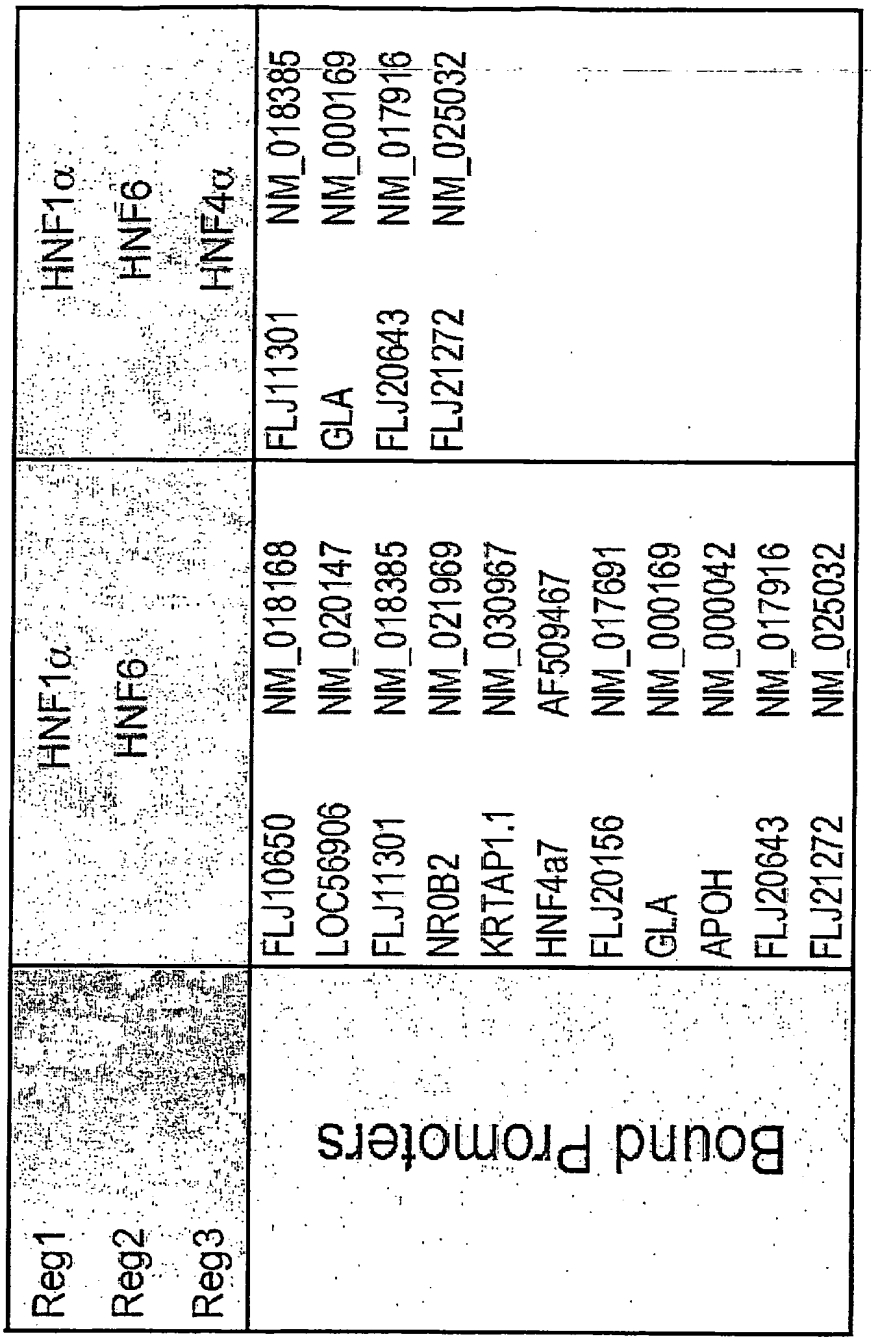

Reg1
Reg2
Reg3
HNF1α
HNF6
HNF4α
HNF1α
HNF6
Bound Promoters
FLJ10650
LOC56906
FLJ11301
NR0B2
KRTAP1.1
HNF4a7
FLJ20156
GLA
APOH
FLJ20643
FLJ21272
NM_018168
NM_020147
NM_018385
NM_021969
NM_030967
AF509467
NM_017691
NM_000169
NM_000042
NM_017916
NM_025032
FLJ11301
GLA
FLJ20643
FLJ21272
NM_018385
NM_000169
NM_017916
NM_025032

Fig. 23B

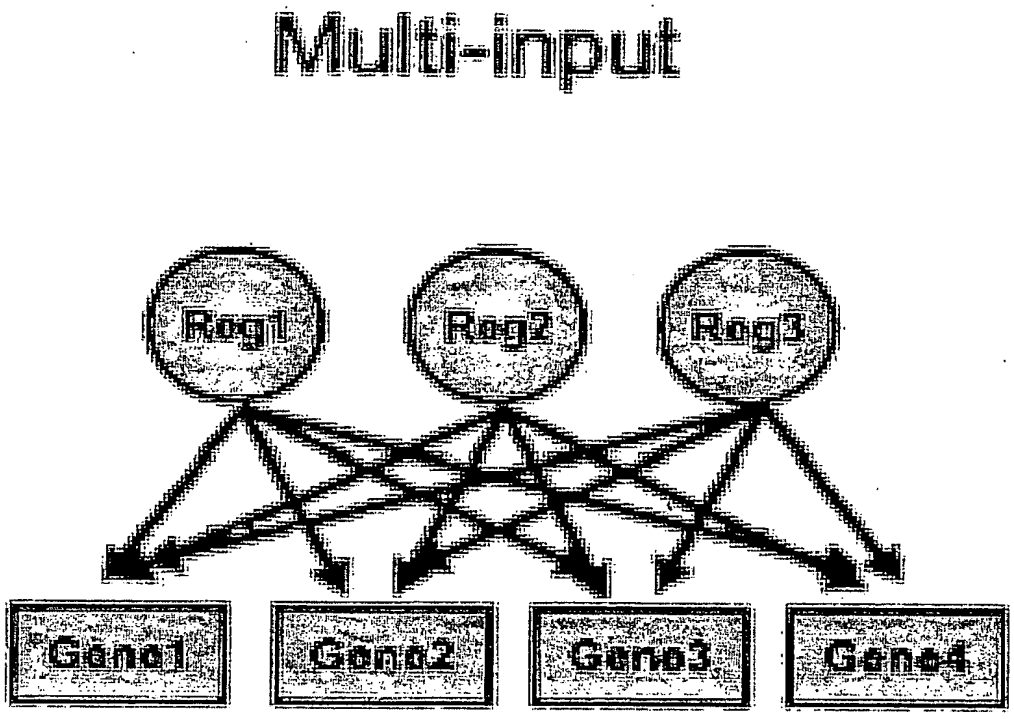
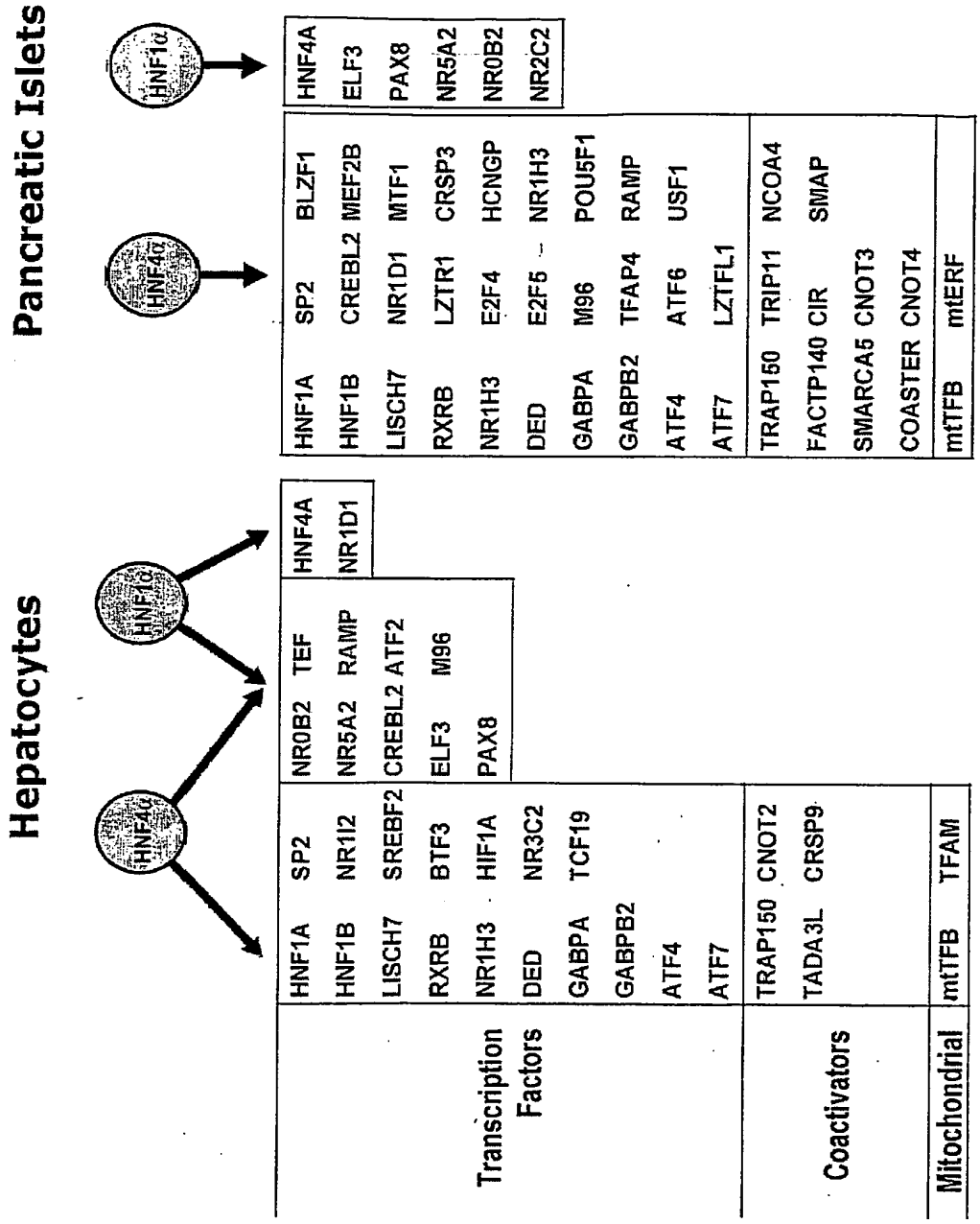


Fig. 24


Hepatocytes
Pancreatic Islets
HNF4α
HNF1α
HNF4α
HNF1α
Transcription Factors
Coactivators
Mitochondrial
HNF1A SP2 NR0B2 TEF HNF4A
HNF1B NR1I2 NR5A2 RAMP NR1D1
LISCH7 SREBF2 CREBL2 ATF2
RXRB BTF3 ELF3 M96
NR1H3 HIF1A PAX8
DED NR3C2
GABPA TCF19
GABPB2
ATF4
ATF7
TRAP150 CNOT2
TADA3L CRSP9
mtTFB TFAM
HNF1A SP2 BLZF1
HNF1B CREBL2 MEF2B
LISCH7 NR1D1 MTF1
RXRB LZTR1 CRSP3
NR1H3 E2F4 HCNGP
DED E2F5 NR1H3
GABPA M96 POU5F1
GABPB2 TFAP4 RAMP
ATF4 ATF6 USF1
ATF7 LZTFL1
TRAP150 TRIP11 NCOA4
FACTP140 CIR SMAP
SMARCA5 CNOT3
COASTER CNOT4
mtTFB mtERF
HNF4A
ELF3
PAX8
NR5A2
NR0B2
NR2C2